Discussion paper

What limits the powerful in imposing the morality of their authority?

BY
Øivind Schøyen

Norges
Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

What limits the powerful in imposing the morality of their authority?

Øivind Schøyen, Norwegian School of Economics[1]

Abstract

This paper models a game between an authority, seeking to implement its preferred morality, and a parental generation, seeking to socialize a younger generation into the their own morality. The authority chooses a coercion level for adhering to the non-state morality, whereupon the parental generation chooses whether to insurrect and, if not, how much to invest in socialization. The novel feature of this paper is that we formalize and explore the consequences of an intrinsic negative reaction to coercion: coercion resentment. The key result is to show the necessary micro-level assumptions for an inefficient interval of coercion that can account for authorities choosing to restrain their use of coercion. Furthermore, the paper characterizes the socialization and insurrection preferences needed for the long run equilibrium to be path dependent. Two historical periods are presented through the lens of the model: the Counter-Reformation in early modern France and the Holy Roman Empire (1517-1685) and the Soviet Secularization project (1922-1991).

Keywords: Moral persistence, Political legitimacy

JEL classification: D02 D10 D82 N30 N40 P16

[1]Department of Economics, NHH Norwegian School of Economics, Bergen, Norway, oivind.schoyen@nhh.no, October 6, 2017.

1. Introduction

Polities generally seek to have legitimacy; that is to rule in alignment with the internalized moralities of the population.[2] A crucial dimension of state legitimacy, and the focus of this paper, is whether the values of the polity, on which it builds its formal institutions and moral right to rule, are aligned with the moralities of its population.[3] One way of attaining legitimacy is by using extrinsic incentives, generally referred to as coercion. This approach may, however, invoke an intrinsic counteraction, making coercion potentially counterproductive. The paper embeds this micro assumption, referred to as coercion resentment, into an overlapping generations model of moralities where an authority seeks to maximize the prevalence of its preferred morality by using coercion. The model analyzes how the opposing effects of the extrinsic incentives to comply, and the intrinsic incentives to resist, determine the prevalence of the different moralities.

Greif and Tadelis (2010) posed the question "Why do the powerful often fail to promote the morality of their authority?". In other words, what are the mechanisms behind moral persistence in the face of hostile institutional environments? If people would simply choose to internalize the morality that gave them the highest extrinsic utility, moralities, and subsequently group identities, would simply be a function of the institutional environment. This would imply that everyone holds the most

---

[2] A morality can be understood as a vector of beliefs and values which is internalized and embedded in a person; examples are political ideologies, religious or ethnic identities.

[3] Another important dimension is whether the state works to fulfill the values on which it builds its institutions, or whether it serves the interest of the individuals who control the state, commonly referred to as corruption (Nye, 1967).

opportune morality; in most cases, the morality aligned with the ruling regime. Dynamics of moralities such as religious, national, or ideological identities could then be ignored in political economy analysis and, at most be treated as a rigidity. Assuming moralities to be a passive function of extrinsic incentives would, however, poorly account for the persistence of minority identities such as the Jews in Europe, states' investments in costly nation building, and foreign nationals' voluntary participation in perilous group conflicts such as the Spanish Civil War (1936-1939).

These historical instances illustrate that intrinsic reactions play an important role in the dynamics of legitimacy and in state development both in the short and long term. Minority moralities in hostile institutional environments can be remarkably persistent, as demonstrated by the historical evidence presented in Greif and Tadelis (2010) of Jews in Medieval Spain, while other historical examples, such as David Laitin's study of the Russian diaspora in the former USSR (Laitin, 1998), show quick adoption of new beliefs, norm sets and national identities, pointing to a rapid change in internalized values.

This paper's main contribution is to build a micro-founded model decomposing the effect of coercion, aimed at changing moralities, into an extrinsic and intrinsic reaction. The extrinsic reaction is a reduction of group identification as a response to incentives, while the intrinsic reaction is a strengthening of in-group identification and out-group resentment in the group being targeted. The model assumes that attempts to force people to change their moralities will invoke a resentment toward the authority behind this use of force, making certain levels of coercion counterproductive to attaining legitimacy and potentially causing insurrections against the authority.

3

A premise of the model is that authorities seek to maximize legitimacy. To any authority, having a high level of legitimacy is desirable for a number of reasons: as Max Weber argues, it increases the probability of staying in power, it reduces costs and expands the possibility frontier of imposing policy (Greif, 2008), and it increases the willingness for altruistic behavior such as conscription (Levi, 1997), or paying taxes (Levi, 1999). The key motivation of states in building national, ideological or religious identities is to make populations respond in a manner that is emotionally related to the morality represented by the state. This is what makes religious and national identities powerful tools for authorities; the ability of internalized norms to invoke reactions that align the interest of the individual with the perceived interest of imagined national, political or religious communities. Further, a population with homogenous moralities enables central policy making (Tilly, 1992); indeed, services such as law and policing, hinge on and grow out of common sets of norms and values.

In the short term, the most obvious way to gain legitimacy is to take norms and values as given, and rule in accordance with the prevailing majority morality. To authorities in polities with heterogenous moralities, this implies making compromises between moralities where they are incompatible, typically at the cost of reduced legitimacy (Johnson and Koyama, 2013). States might, however, enhance their legitimacy by increasing the portion of the population with internalized norms similar to those on which the state builds its institutions. This can be done either by application of "sticks"; disincentives and coercion, or "carrots"; increasing the incentives of

belonging to the authorities' morality.[4] This article focuses on the "stick" approach, coercion, and how it invokes an intrinsic negative reaction, making it a potentially counterproductive measure.[5]

The model develops necessary assumptions for analyzing the equilibrium coercion level and morality prevalence in the overlapping generations model of Bisin and Verdier (2000, 2001). This equilibrium is given as a function of parental preferences for their child adhering to their morality, and the strength and functional form of coercion resentment. A key result of this analysis is to show that authorities will only restrain their use of coercion when there exists an inefficient interval of coercion, which is shown to imply a non-linear response to coercion. The paper then analyzes the dynamic problem of coercion use when the prevalence of the minority morality determines an insurrection constraint on coercion use to find the dynamically stable equilibrium. The key result from this analysis, is to show that the model will exhibit path-dependency; outcomes depend on the history of the polity, if and only if there are coercion levels that can only be implemented from some initial conditions. We also explore the dynamic property of states in coercion reliance; this is defined as an inability for an authority to decrease coercion, as it will increase minority prevalence thus increasing their insurrection capability and trigger an insurrection.

---

[4]Other measures include increasing socialization and easing communication by creating common standards, i.e., through building of roads, language standardization, common school systems and investing in common symbols.

[5]To the extent that "carrots", i.e., positive incentives, invoke a negative reaction amongst the members of the non-state morality, the analysis generalizes to authorities imposing positive incentives for adhering their morality.

The paper presents anecdotal historical evidence to demonstrate macro level restraints of coercion use. First, we review the Counter-Reformation in Early Modern France (1517-1685) and the Holy Roman Empire (1517-1648). The Early Modern French kings and the Holy Roman emperors built their legitimacy on the Catholic faith. The spread of Protestantism following Luther (1517) posed a direct threat to their program of state consolidation. As a response to this introduction of religious heterogeneity, they embarked on programs of homogenization. We argue that in this period, only unconfrontational or strongly coercive policies where stable over time. This supports the model predictions that authorities restrain coercion use, and that any long-term dynamically stable equilibrium must not give the authority any incentive for gradual increases of coercion. The paper then presents a brief comparative study of European early modernity before and after the Peace in Westphalia (1648) through the lens of the model; it shows how the change of international institutions affects constraints on use of coercion and consequently minority prevalence.

Further, we review evidence from the Soviet (1922-1991) secularization policies towards the Christians and Muslims in the USSR. The Soviet Union sought to increase its legitimacy by increasing the support of communism and diminishing the importance of religion. This secularization project was conducted in a comparatively more cautious way in regions where cultural differences towards the Russians were larger, recognizing the potential counter productiveness of secularization attempts, in line with the proposed micro-mechanism of coercion resentment.

The paper develops as follows: The remainder of section 1 reviews related literature, Section 2 presents the overlapping generations model of Bisin and Verdier (2000, 2001) and Section 3 expands the basic model to include a legitimacy maxi-

6

mizing authority, coercion resentment and an endogenously determined insurrection constraint. Section 4 show how the macro predictions of the model fit the Soviet Secularization project and the Counter-Reformation in early modern France and the Holy Roman Empire. Section 5 concludes the paper and discusses questions that can be investigated in future extensions of the theory. The appendixes contains proofs, and some further analysis and interpretations of the model.

1.1. Related literature

The model's critical micro assumption is coercion resentment; individuals are assumed to react negatively towards the authority as a response to coercion. More specifically, it is assumed that at least some levels of coercion for holding a morality will cause individuals to increase their investment in socialization of this morality as a response. Why individuals act in such a way, can be understood from different strands of literature. Three main perspectives are reciprocity, fulfilling internalized norms and increased investment in social motives to help the group faced with a common threat external to the in-group.

Reciprocity: Coercion resentment can be understood as a group level version of the general trait of reciprocity (Bowles and Gintis, 2011); the tendency to retaliate against hostile actions and reward beneficial actions. The assumed mechanism is that individuals that have internalized the coerced morality and feel that the authority has harmed their group, wish to punish the group associated with the coercion through activities aimed at stopping the authorities' influence.

Salience of fulfilling internalized norms: Coercion resentment might also be understood as increased salience of acting in accordance with internalized norms. The

authority, and indirectly, individuals aligned with the state morality, become a salient enemy of the non-state morality if they appear as having hostile intentions. The need to act in line with the non-authority internalized norms will involve confronting individuals of the state morality and stop the spread of their morality once they are conceived as being a threat. In other words, an individual that has internalized a set of values will receive intrinsic utility from actively deterring the influence of an authority pursuing an agenda of opposing his values, as this will help defend his internalized values.

Social motives: A threat from an external foe increases in-group identification. This finding has a longstanding tradition and has solid empirical support in the social psychology literature (Huddy, Sears and Levy, 2013). As coercion towards the non-state morality increases, the authority will be seen as a threat to the non-state morality group. This increased external threat invokes an emotional reaction which triggers investment in social identity activities for individuals that have internalized the non-state morality. The presence of a threat to the group increases in-group identity and strengthens hostility towards the out-group. The out-group threat effect is documented to increase a number of different group related behaviors, including increased investment in socialization (Huddy, Sears and Levy, 2013).[6] Finally, once coercion is imposed on a morality, defying the coercion and acting according to the coerced morality become costly, and can hence be used as a credible social signal of being intrinsically motivated.

---

[6]Although strengthening of a group identity is theoretically different from a utility loss of children adhering to an opposing morality, the modeling implications are similar for the purpose of this study; a society with two mutually excluding moralities.

The paper draws on classical political science analysis of the state's role in moral dynamics. This literature initially focused on cultural unification into nation states, arguing that the relatively high pre-existing (pre 990 AD) homogeneity of morality in Europe contributed to Europe's relatively rapid state consolidation (Tilly and Ardant, 1975), later focusing on the survival and persistence of minority cultures through mechanisms of cultural resistance (Allardt, 1979; Rokkan, 1999).

The paper relates to four strands of the economics literature: social economics, group conflict, state legitimacy and path dependency in societal outcomes. The model is an expansion of the social economics model by Bisin and Verdier (2000, 2001) where overlapping generations transfer moralities, and the prevalence of each morality is determined by parental investment into socialization. Models in social economics have addressed the role of cultural persistence through differences in socialization investment arising by mechanisms such as oppositional culture (Bisin et al., 2011), bias in education systems (Carvalho and Koyama, 2013) and social signaling of identity (Carvalho, 2013). These findings are supported by empirical findings by Fouka (2015) who finds that US citizens subjected to language barriers on German in US schools following the First World War were less likely to volunteer for military service in the Second World War (WW2). Social economics models have generally not focused on actions of state actors or individuals' relation to a state (Bisin and Verdier, 2010). Following Greif and Tadelis (2010), this paper extends the author's master thesis (Schøyen, 2011) and is novel in making the connection between the policies of the state authority and the prevalence of non-state minority moralities. Greif and Tadelis (2010) introduce an authority that controls the institutional environment to maximize the morality on which it builds its legitimacy, into the Bisin and Verdier

9

(2000, 2001) framework. This paper extends Greif and Tadelis (2010) by letting the agents in the model intrinsically react to coercion. In contrast, the agents in the model of Greif and Tadelis (2010) are static in the sense that they do not intrinsically respond to coercion. The paper also contributes by introducing an endogenous dimension of power; an insurrection constraint on the use of coercion dependent on the prevalence of the non-state morality.

The paper also relates to the literature on ethnic and political violence; especially the understanding of the use of force as a root cause of counter-mobilization in the form of a strategic response (Acemoglu and Wolitzky, 2014), or increased saliency of identity due to group conflict (Sambanis and Shayo, 2013). Acemoglu and Wolitzky (2014) focus on the informational aspects of group conflicts that lead to hostile actions being followed by hostile reaction. They develop a dynamic Bayesian game of sequential aggressive or conciliatory actions between groups, where the driving static is whether agents interpret hostile actions of the opposing groups as the actions of a fundamentally aggressive type, or the actions of a non-aggressive type retaliating. They consider the informational aspect of group conflict, while this paper analyzes group conflict as driven by an intrinsic reaction. Sambanis and Shayo (2013) build a formal model endogenizing the process of identification with an ethnic group. They allow for identification on multiple levels and focus on a social identity equilibrium between groups where saliency determines the level of identification. Both these papers consider group relations and their internal dynamic, while in the model presented here, the agency lies in the state authority and the population responses to the level of coercion.

Further, the paper relates to a growing new literature on state legitimacy. The

role of the state in nation building is formally analyzed in the economics literature by Alesina and Spolaore (2003) and Alesina and Reich (2013), while the Greif and Rubin (2014) study illustrates the need for independent agencies to provide legitimacy to the state. Greif and Rubin (2014) consider how the English Crown's breach with the Catholic Church created a need for a new external agent for the king to legitimize his power, thus increasing the need for the approval of an independent agent such as the parliament. Johnson and Koyama (2013) investigate the relationship between the legitimacy gained by aligning the state with a specific religious belief rather than a compromise between several, and the economic cost of enforcement of that belief. Where these papers focus on different sources of legitimacy and alignment between state and morality, this paper focuses on the use of force, its military constraints, and the intrinsic reaction to the use of force aiming to change moralities.

Finally, this paper relates to recent work by Acemoglu and Robinson (2017) in developing dynamic models where path dependency in societal outcomes arises. Acemoglu and Robinson (2017) develop a model of dynamic contest for power where the state and society sequentially makes costly investment into conflict capital. They find path dependency in the power of the state due to the discouragement effect in competitions; the interaction between incentives to invest and economies of scale in capital. This mechanism leads to a dynamic where either state and society invest in conflict capital to be thus equally matched in power, or, where one of the parties ceases to invest and has no power. This paper models the dynamic of available labor for conflict, i.e., sizes of morality groups, when the coercion level changes the size of groups which determine the ability to coerce without having an insurrection. Path dependency arises as non-linearities in the response to coercion and initial sizes of morality groups determines which coercion levels can be implemented by iteratively

changing the coercion level.

2. A basic model of socialization

Following Bisin and Verdier (2000, 2001), we introduce an overlapping generations model where parents invest in costly socialization to make their child internalize the morality of the parents. First, the basics of the model and the mechanisms of socialization are developed. All results here mirror results from Bisin and Verdier (2001). We then develop assumptions on the parents' utility function and derive its implications.

2.1. The model

The population consists of a continuum of agents who live in two periods, as a child at time t and as a parent at time $t + 1$. Each agent produces one offspring, thus the size of the population remains stable. There are two moralities; $m \in \{a, b\}$. Moralities are mutually exclusive; a portion q_t of the parent population holds morality a at time t, while $1 - q_t$ holds morality b. Moralities are transmitted from one generation to the next through parental socialization from parent to child, or through oblique transmission; the influence of the general population. With probabilities τ^m parental socialization is successful and the child adopts the morality of the parent. With probability $1 - \tau^m$, the parental socialization fails, in which case the child is obliquely socialized, and the offspring will adopt either morality a or morality b with a probability equal to the moralities' prevalence in the population. A child who internalizes morality m is referred to as an m morality child. Let P^{mn} be the

12

probability that an individual of morality m has an n morality offspring.

$$P^{aa} = \tau^a + (1 - \tau^a)q_t \ , \ \ P^{ab} = (1 - \tau^a)(1 - q_t) \tag{1}$$

$$P^{bb} = \tau^b + (1 - \tau^b)(1 - q_t) \ , \ \ P^{ba} = (1 - \tau^b)q_t \tag{2}$$

The portion of the population with morality a at time $t + 1$, q_{t+1}, is then given by:

$$q_{t+1} = q_t P^{aa} + (1 - q_t)P^{ba} = q_t + q_t(1 - q_t)(\tau^a - \tau^b) \tag{3}$$

From (3) it follows that the change in the share of morality a individuals is given by $q_t(1-q_t)(\tau^a - \tau^b)$; the difference in the probability of successful parental socialization times the product of the share of moralities.

Parents choose τ^m to maximize expected utility by balancing the cost of parental socialization, denoted by the function $H(\tau^m)$, and the benefit of a higher probability of successful parental socialization. Let the utility of an m morality parent having an n morality child be denoted u_n^m, then using (1) and (2) we attain the following utility function U^m for parents:

$$U^a = [\tau^a + (1 - \tau^a)q_t]u_a^a + (1 - \tau^a)(1 - q_t)u_a^b - H(\tau^a) \tag{4}$$

$$U^b = [\tau^b + (1 - \tau^b)(1 - q_t)]u_b^b + (1 - \tau^b)q_t u_b^a - H(\tau^b) \tag{5}$$

We now impose some assumptions on the parents' preferences over their child's morality and the cost function of parental socialization. First, we assume that parents prefer their child to have the parents' morality:

Assumption 1. Own morality preference Parents favor their child to have the same morality as themselves: $u_a^a - u_b^a > 0, \ u_b^b - u_a^b > 0$.

Second, the utility loss of having a child internalize the opposing morality is assumed to be symmetric for the two types of parents. Defining $\overline{u}$ as the utility derived from the child having the child's parents' own morality and $\underline{u}$ the utility derived from having the opposing morality, we can write the following assumption:

Assumption 2. Symmetric utility loss of opposing morality Parents of a and b morality have symmetric utility loss for having children of the opposing morality: $u_a^a - u_b^a = u_b^b - u_a^b = \overline{u} - \underline{u} = \Delta u.$

Third, we assume that the cost of socializing the child into the preferred morality $H(\tau^m)$ obeys the Inada conditions:

Assumption 3. Inada assumptions Inada conditions apply to the cost of investment in parental socialization: $H'(\tau^m) \geq 0, H'(0) = 0, \lim_{\tau^m \to 1} H'(\tau^m) = \infty, H''(\tau^m) > 0$.

The first part of Assumption 3 states that the marginal cost is increasing with the probability of success and the second that there is no marginal increase of cost of socialization at no parental socialization, $\tau^m = 0$. The third and fourth parts of Assumption 3 state that the marginal cost approaches infinity as the probability of having a child successfully socialized into the preferred morality approaches certainty and that the increase in marginal cost is strictly increasing in τ^m. The assumption of no increase of cost at $\tau^m = 0$ implies that τ^m will be strictly positive whenever the utility of having successful parental socialization is strictly positive for m morality parents. The assumption that the cost of socialization grows towards infinity implies that there will always be some failed parental socialization leading to oblique

14

socialization. Hence, there will always be some children obliquely socialized into the opposing morality in mixed morality populations. We can now derive the optimal levels of τ^m from (4) and (5), which are given by the first order conditions (FOCs).

$$H'(\tau^a) = (1 - q_t)\Delta u, \ H'(\tau^b) = q_t \Delta u \tag{6}$$

The optimal level is given by the expected marginal benefit of investing into parental socialization being equal to the marginal cost. From Assumption 3, the Inada conditions, and (6) we can establish the following lemma:

Lemma 1. The smallest morality group always invests more in parental socialization; $\tau_t^b \leq \tau_t^a$ if and only if $q_t \leq (1 - q_t)$.

Since the benefit of having a child with the parents' morality is assumed symmetric, a difference in investment must imply a difference in the cost of failed parental socialization. Any difference in the utility of failed socialization arises as the probability of the child obliquely internalizing the preferred morality differs due to different group size. The minority parents have a higher probability of the child internalizing the majority morality obliquely if parental socialization fails, and consequently invest more in socialization; hence Lemma 1.

A steady state equilibrium (SSE) level of q, denoted q^*, is reached when $q_t = q_{t+1}$. It follows from (3) that for $q_t = q_{t+1}$ to be fulfilled, $q_t(1 - q_t)(\tau^a - \tau^b) = 0$ must hold. This is the case for $q_t = q_{t+1} = 0$, $q_t = q_{t+1} = 1$, i.e., single morality populations, or, as will be shown, at the interior SSE where $\tau^a = \tau^b$. In cases of $q^* = 0$ or $q^* = 1$, there will be no utility gain of parental socialization as all individuals in the population will have the same morality, and oblique socialization will lead to the preferred

15

morality of the parent. The single morality equilibrium is, however, unstable towards external shocks; if one parent of another morality enters the population, this parent would choose a very high investment into parental socialization since the probability of the child adopting the desired morality in the case of oblique socialization would be very low. This would be repeated for future generations and consequently, the prevalence of the introduced morality of the minority would grow until the unique interior $q^* = \frac{1}{2}$ is reached.

Lemma 2. There is a unique stable interior SSE at $q^* = \frac{1}{2}$.

The only stable equilibrium is $q^* = \frac{1}{2}$; any initial population with a q different from one or zero will converge towards it. Out of the SSE, the share of the minority morality individuals will grow with time as the smaller morality group invests more in socialization, as stated by Lemma 1, until again $q_t = q_{t+1} = \frac{1}{2}$. The fact that the stable interior is $q^* = \frac{1}{2}$ arises due to Assumption 2; symmetry of preferences. Asymmetrical preferences where an interior SSE exists at $\tau^a = \tau^b$, leads to asymmetrical, i.e., $q^* \neq \frac{1}{2}$, stable SSE.[7]

3. Legitimacy maximizing given coercion resentment and an insurrection constraint

We now extend the model to include an authority which can issue a penalty, referred to as coercion, for adhering to the non-state morality. Further, we make assumptions

---

[7]The assumption of symmetric preferences is made in order to focus on the role of the state rather than on any difference between the desirability of the moralities themselves. The following analysis generalizes for asymmetrical preferences.

on how the agents respond to this coercion and analyze the use of coercion under exogenous and endogenous constraints to which different levels of coercion can be imposed. To focus on the implications of coercion resentment, we follow Greif and Tadelis (2010) in assuming that the authority can impose coercion at zero cost. The results can trivially be extended to a model where coercion is costly to the authority.

3.1. The extended model

There is an authority, β, controlling the state, where a state is defined as a monopoly on the employment of coercion, π, within the territory where the population is situated. This authority builds its legitimacy on the b morality and wishes to maximize its prevalence by imposing coercion for adhering to the a morality. The utility maximization problem of the authority, U^β is:

$$\max_\pi U^\beta = \min_\pi q^*(\pi) \tag{7}$$

To maximize the prevalence of the b morality the authority sets the level of coercion π for adhering to morality a. The level of coercion is assumed unbounded; π is defined over the domain $\pi \in [0, \infty)$, but we assume that the authority is bound by an upper feasibility constraint, π_{max}, on the level of coercion it can impose. Hence we restrict our analysis within the feasible interval $\pi \in [0, \pi_{max}]$.[8]

The coercion level can be interpreted as ranging from low, such as social sanctions or issuing fines for having morality a, to high, such as criminal penalties, and the maximum feasible level, π_{max}, referred to as a gunpoint threat. Including the level

---

[8]Feasibility might reflect either technological constraints in terms of what can be implemented, or an upper limit in terms of what the state apparatus will impose.

of coercion, π, and the resentment towards the b morality caused by coercion, $C(\pi)$, in the utility function of the parents of respectively a and b morality yields:

$$U^a = [\tau^a + (1 - \tau^a)q_t](\overline{u} - \pi) + (1 - \tau^a)(1 - q_t)(\underline{u} - C(\pi)) - H(\tau^a) \qquad (8)$$

$$U^b = [\tau^b + (1 - \tau^b)(1 - q_t)]\overline{u} + (1 - \tau^b)q_t(\underline{u} - \pi) - H(\tau^b) \qquad (9)$$

These two utility functions capture the two following assumptions on how the agents respond to coercion.[9] First, we assume that the utility of having an a morality child is lower when there is coercion:

Assumption 4. Parental empathy for coercion The utility of having an a morality child is $(u_a^m - \pi)$.

Second, we assume coercion resentment, imposing coercion invokes a negative intrinsic reaction among the a, non-state morality, parents; i.e., they will have lower utility of having a b morality child:

Assumption 5. Coercion resentment The utility of an a parent of having a b morality child is $(u_b^a - C(\pi))$.

Note that the reduced utility from the introduction of coercion of having an a morality child is the same for parents of both moralities. This reflects the fact that a

---

[9]At $q_t \in \{0,1\}$ parents will be indifferent between successful parental socialization or oblique socialization. Since the cost of investing some infinitesimal amount into socialization or investing nothing $\tau^m = 0$ is equal they will be indifferent between these two outcomes. We impose that, $\tau^m = 0$, for $q_t \in \{0,1\}$.

morality children face an extrinsic cost, while the coercion resentment cost of having a b morality child, an intrinsic loss of utility, is limited to affect the non-state morality parents. We now derive the optimal levels of τ^m from (8) and (9), which are given by the FOCs:

$$(1 - q_t)(\Delta u - (\pi - C(\pi))) = H'(\tau^a) \tag{10}$$

$$q_t(\Delta u + \pi) = H'(\tau^b) \tag{11}$$

Comparing (10) and (11) with the FOCs in the baseline model, (6), we see that the b morality parent increases the socialization investment as π incurs a more severe utility loss if parental socialization fails and the child obliquely internalizes a morality. For a morality parents, coercion introduces two opposing effects; they have an incentive to reduce their investment as having an a morality child becomes less extrinsically beneficial, and an incentive to increase investment as b morality becomes less intrinsically beneficial as a result of coercion resentment. Without assuming a functional form on the coercion resentment function, we cannot say which effects dominate at which coercion levels.

As developed in the basic model of socialization, we see from (3) that a necessary condition for a stable interior SSE level is equal levels of investment, τ^m, in socialization for a and b morality parents. If parents invest equally much in socialization, they have equal marginal costs, $H'(\tau^a) = H'(\tau^b)$, hence we can use (10) and (11) and establish the following lemma:

Lemma 3. For all pairs of $\{\pi, \Delta u\}$ two exterior SSEs exist. For some, but not all, pairs of $\{\pi, \Delta u\}$ a unique stable interior SSE exists, given by $q^*(\pi) = \frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}$.

This result is a basic extension of Proposition 1 in Greif and Tadelis (2010). The stationarity properties of (3) imply that the population will always converge to its SSE value. The Inada assumptions on investment in socialization imply positive investment in socialization for a π corresponding to an internal SSE, hence the population will not reach any exterior solution in the convergence process as long as the SSE is internal. We define an initial interior SSE as some $q^*(\pi_0) \in (0,1)$, where π_0 is some initial coercion level $\pi_0 \in [0, \pi_{max}] : q^*(\pi_0) \in (0,1)$ and establish this as the following lemma:

Lemma 4. Imposing a coercion level π' corresponding to an internal SSE, $q^*(\pi') \in (0,1)$, from an initial interior SSE $q^*(\pi_0)$ will make q converge to $q^*(\pi')$.

This result is a basic extension of Proposition 1 and Proposition 2 in Bisin and Verdier (2001). To illustrate the dynamics of the model, let us assume that at time t the coercion level is $\underline{\pi}$ and the population is in an interior SSE with $q^*(\underline{\pi})$. Assume that the π changes at $t+1$ to $\overline{\pi}$, where $\overline{\pi} > \underline{\pi}$, and that the net effect of coercion for a morality parents, $(\pi - C(\pi))$, is sufficiently increasing in the interval $[\underline{\pi}, \overline{\pi}]$ such that $q^{*\prime}(\pi) < 0$.[10] In $t+1$, q remains unchanged but investment in socialization changes; the a parents will now invest less in socialization as they have a net lower utility of having a morality children, while the b parents will invest more in socialization as the outcome of unsuccessful parental socialization, having an a morality child, is less desirable to them. Socialization efforts now differ and q drops to $q_{t+2} < q^*(\overline{\pi})$ for the first generation presiding the change in π. At time $t+2$, parents will make

---

[10]This implies the average $C'(\pi)$ is less than two in the interval, as can be seen from (31) in Appendix 2.

the socialization investment decision with q_{t+2}, which is strictly smaller than q_{t+1}. Hence, the a morality parents will face a higher probability of their offspring having a b morality through oblique socialization and consequently increase their parental socialization. The level of the minority morality q_t will converge toward $q^*(\overline{\pi})$ until the SSE condition from (3) of $\tau^a = \tau^b$, i.e., equal investment in parental socialization, is restored at the SSE with $q^*(\overline{\pi})$.

Imposing a coercion level that does not correspond to an internal SSE, must imply that π such the one of the morality groups will cease to invest in socialization. This will lead to $q^*(\pi)$ reaching the external SSE without the morality group that ceases to socialize their children within one generation.

We now proceed to discuss the coercion resentment function. The form of the coercion resentment function can be understood as a normalization of the effect of coercion resentment relative to the effect of coercion normalized to a unit scale, i.e., assumed to be simply π. Thus, discussion of the net effect of coercion for the a morality parents can be centered around the coercion resentment function, $C(\pi)$. First, some fairly unrestrictive functional form assumptions are made on $C(\pi)$:

$C(\pi)$ is a function of the C^2 class, it is $C(0) \geq 0$ and it has $C'(\pi) > 0$, over the domain $[0, \pi_{max}]$.

(12)

C^2 is the class of functions for which first and second derivatives are continuously defined over the entire domain of the function.

The coercion resentment function is assumed to be S-shaped.[11] The convex part of the coercion function captures the idea that there is an increasing marginal emotional response to an increase in π for the initial levels of coercion. As the authority increases π, it goes from being perceived as a representative of the b morality, which favors and endorses the b morality, to being perceived as an enemy of the a morality individuals, with aggressive intentions of reducing the prevalence of the a morality. The concave interval reflects the increase of this response diminishes beyond some point; as the intentions of the authority have become clear, higher levels of coercion cause a smaller increase in resentment. We define a point $\hat{\pi}$ in the open interval, $\hat{\pi} \in (0, \pi_{max})$ and assume that:

$$C''(\pi) = \begin{cases} > 0 & \text{for} \quad \pi \in [0, \hat{\pi}) \\ = 0 & \text{for} \quad \hat{\pi} \\ < 0 & \text{for} \quad \pi \in (\hat{\pi}, \pi_{max}] \end{cases} \tag{13}$$

Further we make the following assumption on the $C(\pi)$ function:

Assumption 6. Varying coercion resentment The marginal utility loss due to coercion resentment approaches zero at the beginning and at end of $[0, \pi_{max}]$; $\lim_{\pi \to 0} C'(\pi) = 0, \lim_{\pi \to \pi_{max}} C'(\pi) = 0$, and is strictly larger than two at least one point, $\pi' \in (0, \pi_{max}); C'(\pi') > 2$.

We define a coercion level π' as marginally effective if $q^{*'}(\pi') < 0$. Using the assumptions on $C(\pi)$ in (12), (13) and Assumption 6 we can develop the following lemma on the overall effect of coercion:

---

[11] As is shown in Appendix 2, linear, convex or concave coercion resentment functions have trivial and unique optimums.

Lemma 5. Coercion is marginally effective at the beginning and at the end of $[0, \pi_{max}]$, and there is at least one level of coercion, $\hat{\pi}$, that is strictly marginally ineffective; $q^{*\prime}(\hat{\pi}) > 0$.

The existence of a level of coercion that is strictly marginally ineffective, preceded and followed by marginally effective levels of coercion, is a crucial assumption on which the following results rest; variation in the marginal efficiency of coercion. With no variation in the marginal effectiveness of coercion, i.e., if all levels of coercion in $[0, \pi_{max}]$ were marginally effective or were strictly marginally ineffective, the result would be trivial; the authority would either always apply the maximum level of coercion or never apply any coercion at all.

Whenever $q^{*}(\pi)$ is strictly positive for all $\pi \in [0, \pi_{max}]$, the assumptions that $C(\pi)$ is monotonically increasing in π and has a continuous second derivative, imply that $q^{*}(\pi)$ will always have a unique infimum in the open convex part of $C(\pi)$, $(0, \hat{\pi})$ where $q^{*\prime}(\pi) = 0$. We denote the coercion level giving this infimum as $\pi_{\underline{q}} \in (0, \hat{\pi})$; and refer to it as a non-confrontational level of coercion. Further, we denote $\pi_{\underline{q}}^{e}$ to be the first coercion level larger than $\pi_{\underline{q}}$ that has $q^{*}(\pi)$ equal to the unconfrontational level:

$$\pi_{\underline{q}}^{e} \text{ is defined as a coercion level such that } \pi_{\underline{q}} < \pi_{\underline{q}}^{e} \text{ and } q^{*}(\pi_{\underline{q}}^{e}) \equiv q^{*}(\pi_{\underline{q}}). \quad (14)$$

$\pi_{\underline{q}}^{e}$ will only be defined for functional forms where $C(\pi)$ is sufficiently concave in $(\hat{\pi}, \pi_{max}]$. There will always be a unique supermum value of $q^{*}(\pi)$ in the concave part of $C(\pi)$, we denote this level as $\pi_{\bar{q}} \in (\hat{\pi}, \pi_{max}]$. Applying Lemma 5 on the effect of coercion, the assumptions placed on $C(\pi)$ in (12) and (13) and Lemmas 3 and 4

23

on interior SSE, we can develop Lemma 6 on the functional form of $q^*(\pi)$:

Lemma 6. $q^*(\pi)$ is characterized by the following properties:

I) A unique global or local maximum($\pi_{\overline{q}}$) and a unique global minimum($\pi_{\underline{q}}$).

or

II) A unique global or local maximum($\pi_{\overline{q}}$), a local minimum($\pi_{\underline{q}}$), and a global, potentially unique, minimum ($\pi' \in [\pi_{\underline{q}}^e, \pi_{max}]$).

or

III) A global minimum ($\pi'' \in (0, \hat{\pi})$, where $q^*(\pi'') = 0$).

In addition, there will always be a local or unique global maximum at $q^*(0) = \frac{1}{2}$. [12]

The properties of $q^*(\pi)$ are dependent on the size of the utility loss for parents of having children in a differing morality, Δu, and on the strength of the coercion resentment relative to the intrinsic effect of coercion. Class III) applies when Δu is sufficiently small and coercion resentment is sufficiently weak such that a coercion level $\pi'' \leq \pi_{\underline{q}}$ gives $q^*(\pi'') = 0$. If $q^*(\pi_{\underline{q}}) > 0$, then either class II) or class I) applies, depending on the concavity of $C(\pi)$ in $(\hat{\pi}, \pi_{max}]$; if $C(\pi)$ is sufficiently concave such that $q^*(\pi_{\underline{q}}) > q^*(\pi_{max})$ then class II) applies, if not, then $\pi_{\underline{q}}$ is a global minimum, and I) applies. Note that class I) is qualitatively similar to a convex $C(\pi)$; it has a unique non-zero minimum $q^*(\pi)$ value, class III) is qualitatively similar to a linear coercion resentment function, i.e., $C(\pi) = K_0 + K_1\pi$, with $K_1 > 2$, while II) is qualitatively non-convex. Figure 1. illustrates the three possible classes of $q^*(\pi)$.

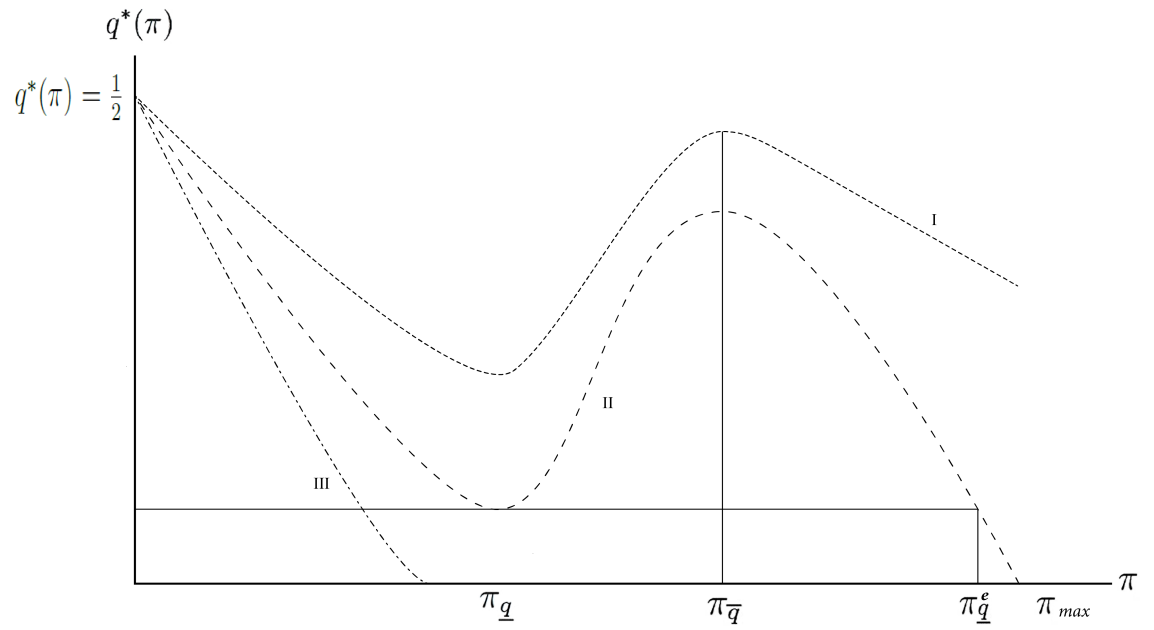[12]When $q^*(\pi)$ is characterized by III) it may also have a unique global or local maximum ($\pi_{\overline{q}}$).

Figure 1: Three examples of $q^*(\pi)$ from $\pi_0 = 0$ constructed using $C(\pi) = \tan^{-1} \pi$.

We now analyze the optimal level of coercion for an authority minimizing the share of individuals with *a* morality in the SSE; $q^*(\pi)$. We assume that the authority chooses π from an initial interior SSE, $q^*(\pi_0)$. First, we discuss the model without any constraint on the use of coercion within $[0, \pi_{max}]$. Second, we discuss the optimal coercion levels under an exogenously given constraint on coercion, $\rho \in (0, \pi_{max})$. Finally, we discuss the model under an endogenously given constraint $\rho(q^*(\pi)) \in [0, \pi_{max}]$. The no constraint analysis is done to analyze how different functional forms on $C(\pi)$ within $[0, \pi_{max}]$ change the optimum. The introduction of constraints within the $[0, \pi_{max}]$ interval reflects strategic constraints which are introduced to show how a change of strategic constraints within one interval changes the optimal level of coercion. Since the initial coercion level, π_0, will not influence the optimal choice of the authority for the no or exogenous constraint cases, it is omitted from the analysis of these cases and only introduced for the endogenous constraint analysis.

Note that in this analysis we do not restrict the optimal choice of π to lead to an interior solution; if the authority can impose a π' that corresponds to the exterior solution, $q^*(\pi') = 0$, it will do so.[13]

No constraint

---

[13]The utility function in (7) implies that whenever a set, i.e., multiple, (π'', π''') corresponds to $q^*(\pi) = 0$, the authority will be indifferent to which $\pi \in (\pi'', \pi''')$ to impose; by institution we refer to the imposed level as the lowest π that attains $q^*(\pi) = 0$. Once the population is in an exterior SSE, q^* will no longer be a function of π and the model is silent on which π the authority will impose.

As established in Lemma 4 and subsequent discussion, an authority in $q^*(\pi_0)$ can choose any feasible π' and will always converge to the corresponding $q^*(\pi') \in [0,1]$. Applying Lemma 3 and Lemma 4 to the utility function in (7), the maximization problem for an authority is given by:

$$\max_{\pi} U^\beta = \min_{\pi}[\frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}] = \min_{\pi} q^*(\pi) \qquad (15)$$

The optimal coercion level under no constraints on the use of coercion, π^{NC}, is determined by the properties of $q^*(\pi)$, given by $\{\Delta u, C(\pi)\}$. Consequently the results follow directly from Lemma 6. For sufficiently strong coercion resentment and large Δu, class I) applies. For class I), the imposed level of coercion will be $\pi^{NC} = \pi_{\underline{q}}$; as $q^*(\pi_{\underline{q}})$ is strictly smaller than $q^*(\pi_{max})$, and will consequently be preferred by the authority. Hence, when coercion resentment is sufficiently strong, such that the $q^*(\pi)$ is at its minimum for unconfrontational levels of coercion, the authority will not impose its maximum level of coercion, even when it is able to do so. The population will remain in an interior solution with the presence of both morality types with an unconfrontational level of coercion.

If class II) applies for $q^*(\pi)$, the optimal level of coercion will be $\pi^{NC} \in (\pi_{\underline{q}}^e, \pi_{max}]$ if any π' gives $q^*(\pi') = 0$. Otherwise, $\pi^{NC} = \pi_{max}$ will be the optimal level of coercion and the population will be stable in an interior solution at $q^*(\pi_{max})$. If class III) applies to $q^*(\pi)$, Δu is sufficiently low and coercion resentment is sufficiently weak so that the authority can impose a coercion level which is lower than the unconfrontational level and attain $q^* = 0$. We sum up the no constraint analysis in Proposition 1:

27

Proposition 1. Coercion use under no constraint Let π' denote a level of coercion such that $q^*(\pi') = 0$. The optimal level of coercion under no constraint, π^{NC}, will be as follows for the different classes of $q^*(\pi)$:

I) $\pi^{NC} = \pi_{\underline{q}}$.

II) $\pi^{NC} = \pi' \in (\pi_{\underline{q}}^e, \pi_{max})$, if no π' is defined then $\pi^{NC} = \pi_{max}$.

III) $\pi^{NC} = \pi' < \pi_{\underline{q}}$, where π' is always defined.

Referring to an authority which imposes a coercion level that is strictly lower than its highest implementable level towards a non-zero a morality group as exhibiting restraint, we can establish the Corollary of Proposition 1:

Corollary of Proposition 1. Restraints under no constraint An authority facing no constraint on coercion will only restrain its use of coercion when SSE $q^*(\pi)$, is of class I).

<center>Exogenous constraint</center>

Several factors external to the model can constrain the use of coercion for an authority; the authority might recognize constitutional legal rights, there might be institutionalized rights causing constraints on what π the state apparatus can issue, or surpassing a coercion threshold might trigger an intervention by foreign powers. To analyze optimal use of coercion when the authority's ability to impose coercion is limited, π^{EC}, an exogenous constraint $\rho \in (0, \pi_{max})$ is introduced. We assume an initial coercion level $\pi_0 \in [0, \rho]$ from which any $[q^*(0), q^*(\rho)]$ can be reached. In addition to analyzing what is the optimal π^{EC}, we also focus on when the constraint

<center>28</center>

will be binding at the optimal coercion level.[14]

Including a constraint on the use of coercion leaves an authority with the following optimization problem:

$$\max_{\pi} U^{\beta} = \min_{\pi} \left[\frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)} \right] \text{ s. t. } \pi \leq \rho \tag{16}$$

Trivially, an exogenous constraint $\rho \in (0, \pi_{max})$ affects the optimal level of coercion π^{EC} if, and only if, it is strictly smaller than the optimal adjustment under no constraint, $\rho < \pi^{NC}$. Noting this, we can develop the following proposition on the optimal level of coercion, π^{EC}, for an authority facing a constraint on the use of coercion:

Proposition 2. Coercion use under an exogenous constraint If a constraint affects the coercion use under an exogenous constraint, $\rho \leq \pi^{NC}$, and $\rho \neq \pi_{\underline{q}}^e$, it holds that:

(i) $\pi^{EC} = \rho$ if and only if $\rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.

(ii) $\pi^{EC} = \pi_{\underline{q}} < \rho$ if and only if $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.[15]

The result shows that constraints in $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$, an interval of coercion which an authority will always find it undesirable to impose, referred to as the inefficient interval of coercion, lead to an imposed level of coercion $\pi^{EC} = \pi_{\underline{q}}$ with a constraint that is non-binding in optimum; $\pi^{EC} < \rho$. Further, if ρ changes from within $\rho' \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$

---

[14]Binding constraints might change the level and the saliency of conflict between morality groups. Further, it will presumably be easier to empirically observe binding constraints, such as explicit threats of intervention and emerging population movements, than unbinding and latent constraints which might be in the form of unrealized outcomes anticipated by an authority.

[15]With $\rho = \pi_{\underline{q}}^e$ the authority is indifferent between imposing $\pi_{\underline{q}}$ and $\pi_{\underline{q}}^e$.

to some level $\rho'' > \pi_{\underline{q}}^e$, the coercion level will jump discontinuously from $\pi^{EC} = \pi_{\underline{q}}$ to $\pi^{EC} = \rho''$.[16]

Considering Lemma 6, we see that the inefficient interval is only defined for functional form II); we use this together with Proposition 2 to develop the following proposition:

Corollary of Proposition 2. Restraints under constraint An authority will restrain its use of coercion as a response to a constraint if and only if the $q^*(\pi)$ is of class II) and the constraint is in the inefficient interval of coercion; $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.

Hence; a legitimacy maximizing authority rationally restraining its level of coercion as a response to a constraint, must imply a non-convex response to coercion. In this model, it also implies that the imposed coercion level π^{EC} being equal to the unconfrontational level of coercion $\pi_{\underline{q}}$.

Endogenous insurrection constraints

We now analyze the model, assuming an endogenous insurrection constraint on the use of coercion, dependent on the initial prevalence of a morality, $q^*(\pi_0)$. For tractability, the insurrection constraint $\rho(\cdot)$, is assumed to be dependent on $q^*(\pi)$

---

[16]Proposition 2 has relevant implications for the policy problem of an external agency setting a constraint ρ to limit an authorities' use of coercion when $q^*(\pi)$ discussed in Appendix 3.1.

rather than on q_t; $\rho(q^*(\pi))$.[17]

We first define the insurrection constraint, before we show that the solution to the static optimizing problem of setting π from an initial π_0 is not necessarily an equilibrium if the authority has the opportunity to reset π in the new SSE, $q^*(\pi)$.[18] To address this issue, we develop a formal definition of the set of implementable coercion levels, $\mathbf{S}_{\pi_0}$, for an authority with a given initial condition π_0. To find which of the implementable levels of coercion will be an equilibrium outcome, we develop a notion of dynamically stable equilibria, characterized by the authority not having an incentive to change π^{IC} if this was the initial coercion level; $\pi_0 = \pi^{IC}$.[19] Finally, we investigate which coercion levels characterize dynamically stable equilibria and show how the model may display path dependence, i.e., different initial conditions may give different equilibria.

The endogenous insurrection constraint is defined as the highest coercion level for which the minority has negative excepted utility of committing an insurrection. The insurrection constraint function $\rho(q^*(\pi_0))$ defines the maximal coercion level that can be implemented for some initial state $q^*(\pi_0)$ without the a morality committing an

---

[17]A description of which conditions are needed for equivalence between solving the authorities' optimization problem constrained by an insurrection constraint dependent on the SSE $q^*(\pi)$, $\rho(q^*(\pi))$ or by a constraint dependent on q_t, is included in Appendix 4.1.

[18]Note that since Lemma 5 implies convergence from an internal to another internal SSE, we cannot say that the authority can reset π once q^t reaches $q^*(\pi)$. The issue can be solved by assuming that the authority can reset π once q^t is within some infinitesimal interval ϵ of $q^*(\pi)$. This is, however, not necessary to address which stable equilibria exist and are reachable within $t \in [0, \infty)$, which is the subject of this model, hence we omit this complication.

[19]A standard definition of stability in dynamic games (Petrosyan, 2016).

insurrection. Note that there is made no explicit link between the insurrection deci-sion and coercion resentment; the private decision processes of how much to invest into socialization may be very different to the public decision process for a morality group to commit an insurrection. There is no specified outcome for an insurrection; since we assume the authority to set π in order to avoid an insurrection, we implic-itly assume that the authority must find the insurrection outcome to be worse than being able to reset a π, satisfying the constraint. Implicitly we also assume that the minority might avoid or reduce coercion given a successful insurrection. We first impose the following assumption on the insurrection constraint:

Assumption 7. Monotonically increasing insurrection constraint The insur-rection constraint, $\rho(q^*(\pi))$, is a continuous mapping from $q^*(\pi) \in (0,1)$ to $[0, \pi_{max}]$. It is monotonically decreasing in $q^*(\pi)$ and has a continuous first derivative.[20]

We insert the endogenous insurrection constraint into (15) to attain the authori-tys' static optimization problem with an endogenous insurrection constraint:

$$\max_{\pi} U^\beta = \min_{\pi} [\frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}] = \min_{\pi} q^*(\pi) \text{ s.t. } \pi \leq \rho(q^*(\pi_0)) \qquad (17)$$

Unless the authority can set π only once, and is unable to subsequently readjust its π, the solution to (17) is not necessarily a dynamically stable equilibrium. Since the insurrection constraint is dependent on $q^*(\pi)$, choosing the optimal $\pi = \pi'$ from

---

[20]We discuss interpretations of the insurrection constraint in Appendix 3.2.

an initial condition $q^*(\pi_0)$ may imply that the new insurrection constraint is less binding, $\rho(q^*(\pi_0)) < \rho(q^*(\pi'))$. Hence, the π' solving (17) may be dynamically unstable in the sense that the authority may have an incentive to set a new $\pi'' > \pi'$ in order to attain a lower SSE, $q^*(\pi'') < q^*(\pi')$.

To find which dynamically stable coercion level an authority *will* implement, we first develop a formal notion of which coercion levels an authority *can* implement if it has the opportunity to reset π an infinite number of times; $\mathbf{S}_{\pi_0}$. We first define the set of sustainable coercion levels, $\mathbf{S}_\Pi \equiv \{\pi : \pi \le \rho(q^*(\pi))\}$; these are the levels of coercion that, at their corresponding SSE level, do not breach the insurrection constraint. Since it is not necessarily the case that all $\pi \in \mathbf{S}_\Pi$ are *implementable* from a given initial condition, π_0, $\mathbf{S}_{\pi_0}$ is a subset of the set of sustainable levels; $\mathbf{S}_{\pi_0} \subseteq \mathbf{S}_\Pi$. We formally define the set of implementable coercion levels, $\mathbf{S}_{\pi_0}$, from an initial condition π_0 as follows:

Definition of the set of implementable coercion levels: A coercion level π' is in the set of implementable coercion levels $\mathbf{S}_{\pi_0}$ if and only if there exists a finite sequence, $\{\pi_n\}_0^N \equiv \{\pi_0, \pi_1, \pi_2, ..., \pi_N\}$ where $N \in [0, \infty)$, with $\pi^N = \pi'$, that satisfies the two criteria:

1. Every coercion level in $\{\pi_n\}_1^N$ is implementable from its previous value; $\pi_n \le \rho(q^*(\pi_{n-1}))$ for all $n = 1, 2...N$.

2. Every coercion level in $\{\pi_n\}_1^N$ is sustainable; $\pi_n \in \mathbf{S}_\Pi$ for all $n = 1, 2...N$.

The set of SSE levels corresponding to the set of implementable coercion levels is

denoted $\mathbf{Q}_{\pi_0} \equiv \{q^*(\pi) : \pi \in \mathbf{S}_{\pi_0}\}$. $\mathbf{S}_{\pi_0}$ and $\mathbf{Q}_{\pi_0}$ will be non-empty for any π_0.[21][22]

To further study $\mathbf{S}_{\pi_0}$, we develop the composite function, $\hat{\rho}(\pi) \equiv \rho(\cdot) \circ q^*(\pi)$, as the composite of the insurrection constraint $\rho(q^*(\pi))$ and $q^*(\pi) \in (0,1)$, i.e., a value of $\hat{\rho}(\pi')$ is the insurrection constraint at the SSE corresponding to π'; $q^*(\pi')$.[23] The functional form of $\hat{\rho}(\pi)$ will determine the properties of $\mathbf{S}_{\pi_0}$ and will be determined by the form of $q^*(\pi)$ in conjunction with the form of the insurrection constraint $\rho(\cdot)$. Since $\rho'(q^*(\pi))$ is assumed to be monotonically increasing in $q^*(\pi)$, the functions $\hat{\rho}(\pi)$ and $q^*(\pi)$ will always have derivatives of equal sign.[24] Plotting an illustration of the $\hat{\rho}(\pi)$ function exemplified by a convex $\rho(\cdot)$ function is done in Figure 2.

Not all implementable coercion levels are sustainable; if a coercion level π' increases the SSE from $q^*(\pi_0)$, it will decrease the insurrection threshold of the a morality group and might lead to an insurrection at $q^*(\pi')$. Hence, a coercion level, π', might be implementable, $\hat{\rho}(\pi_0) > \pi'$, lower than an authority's initial condition, $\pi' < \pi_0$, but still be unsustainably *low*. A coercion level may also be implementable but unsustainably *high*. If an authority was to implement an implementable but unsustainable

---

[21]By assumption, the initial condition corresponds to a sustainable level of coercion; $\pi_0 \leq \rho(q^*(\pi_0))$.

[22]Implicitly we here assume that π can only be reset once a $q^*(\pi)$ is reached. A discussion of the set of implementable coercion levels where the authority can reset π at any t is included in Appendix 4.2.

[23]Since $\rho(\cdot)$ is a continuous function mapping from $q^*(\pi) \in (0,1)$ to $\pi \in [0, \pi_{max}]$, and the function $q^*(\pi) \in (0,1)$ is continuous function mapping from $[0, \pi_{max}]$ to $q^*(\pi) \in (0,1)$, the composite of the two, $\hat{\rho}(\pi)$ is a continuous function mapping from $[0, \pi_{max}]$ to $[0, \pi_{max}]$.

[24]This trivially holds since $\hat{\rho}(\pi) \equiv \rho(q^*(\pi))$ and Assumption 7 statement $\rho(\cdot)' > 0$ for all $q^*(\pi) \in (0,1)$ implies that when $q^*(\pi)' > 0$ then $\hat{\rho}(\cdot)' > 0$.
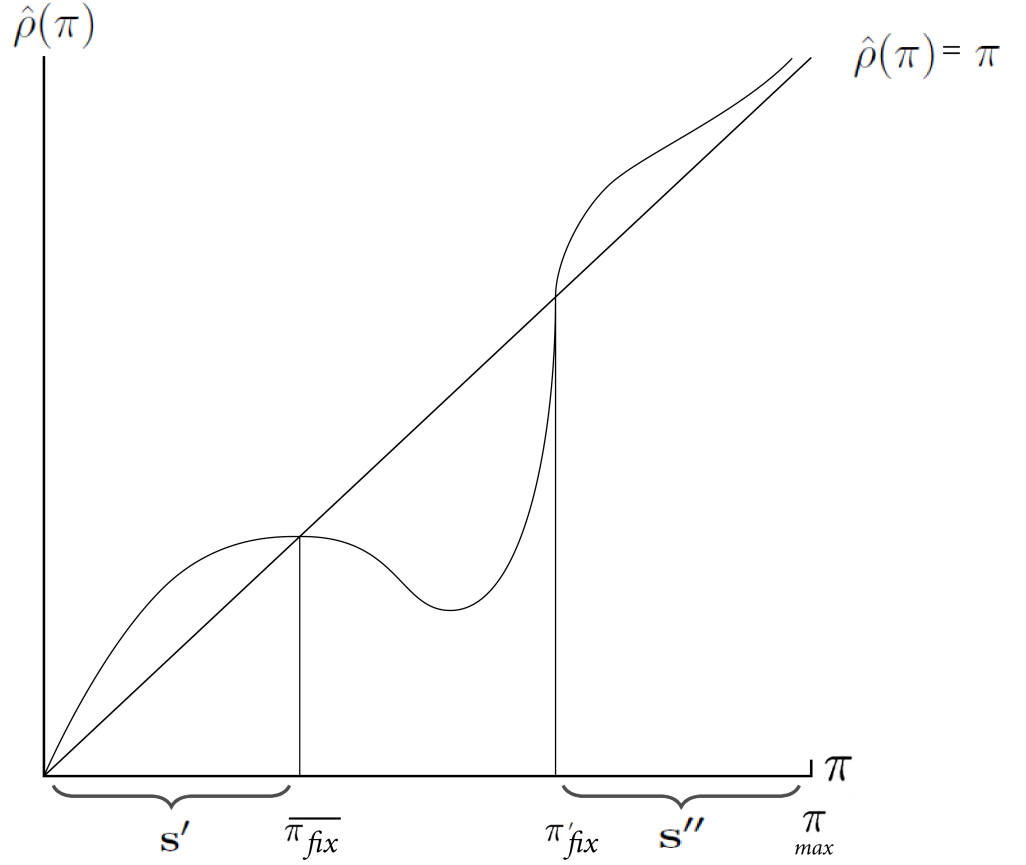
Figure 2: The curved line is an example of $\hat{\rho}(\pi)$ while the forty-five-degree line is the fix-point-line. Any $\pi_0 \leq \overline{\pi}_{fix}$ will give $\mathbf{S}_{\pi_0} = \mathbf{s}' = [0, \overline{\pi}_{fix}]$, while a $\pi_0 \geq \pi'_{fix}$ will give $\mathbf{S}_{\pi_0} = \{\mathbf{s}', \mathbf{s}''\} = \{[0, \overline{\pi}_{fix}], [\pi'_{fix}, \pi_{max}]\}$.

coercion level π', the state would remain stable at the time of implementation, but have an insurrection once q converges to its $q^*(\pi')$ level, at which the size of the a morality group is large enough that they will choose to commit an insurrection at π'. Whenever a coercion level is implementable, $\hat{\rho}(\pi_0) > \pi'$, but also unsustainable $\hat{\rho}(\pi') < \pi'$, it implies that $q^*(\pi)$ is strictly increasing in the interval between π_0 and π'. Hence, an authority minimizing $q^*(\pi)$ will never implement an unsustainable low π in a path towards an π^{IC}; since an unsustainable level of coercion must imply imposing a π increasing the SSE.[25]

Due to the non-linearities in $q^*(\pi)$ there might exist unsustainable levels of coercion π' between the upper and the lower bound of a $\mathbf{S}_{\pi_0}$. We define any sub-set of $\mathbf{S}_{\Pi}$ which is not a union of two disjoint non-empty open sets as $\mathbf{s}$.[26] Since $\hat{\rho}(\pi)$ may have multiple crossings of the fix-point-line, there may be multiple $\mathbf{s}$ sets separated by unsustainable π in a $\mathbf{S}_{\pi_0}$. As $\rho(q^*(\pi))$ is strictly decreasing in $q^*(\pi)$, an authority with $\pi_0 \in \mathbf{s}'$ will always be able to set $\pi' = \hat{\rho}(\pi_0)$ until it reaches an end point of $\mathbf{s}'$, due to the non-linearities in $q^*(\pi)$ this is not necessarily true for $\mathbf{S}_{\pi_0}$ i.e., more sophisticated programs of changes in the coercion level might be needed.

Authorities with composite insurrection constraints such that zero is not a sustainable coercion level, $0 \notin \mathbf{S}_{\Pi}$, are defined to be strongly coercion reliant; they will be reliant on strictly positive levels of coercion to sustain their state, and impose $\pi > 0$

---

[25]An example of an authority that could implement a unsustainably low π, is one seeking to gradually reduce π that knows the direction of $q^{*\prime}(\pi)$ but uncertain of the magnitude.

[26]This definition is adopted from Mendelson (1975).

without any inherent incentive to minimize $q^*(\pi)$.[27] We establish the following as a formal definition of coercion reliance:

Definition of weak and strong coercion reliance: An authority is defined as *strongly coercion reliant* whenever $\hat{\rho}(0) < 0$ and *weakly coercion reliant* whenever there exists unsustainable levels of coercion π' lower than the initial condition; $\pi' < \pi_0 : \pi' \notin \mathbf{S}_{\pi_0}$.

Since $\rho(\cdot)$ is homogenously decreasing in $q^*(\pi)$, a strictly marginally inefficient interval of coercion between intervals of marginally effective π, such as $(\pi_{\underline{q}}, \pi_{\overline{q}})$, must exist for weak coercion reliance to occur independently of strong coercion reliance.[28] We establish this as Proposition 3:

Proposition 3. Weak coercion reliance and coercion inefficiency Weak coercion reliance occurring without strong coercion reliance implies a strictly marginally inefficient interval of coercion between intervals of marginal effective coercion levels.

The bounds of $\mathbf{S}_{\pi_0}$ will either be a coercion level at which the insurrection constraint binds at the $q^*(\pi')$ to which it has converged; i.e., at the line of fix-points, referred to as the fix-point-line, of $\hat{\rho}(\cdot)$ defined as $\pi_{fix} \equiv \{\pi : \hat{\rho}(\pi) = \pi\}$, or the

---

[27] Assuming that a group with superior coercive capability will always commit an insurrection, an authority with coercion reliance, $0 \notin \mathbf{S}_{\pi_0}$, implies that the minority has superior coercive capability at $q^*(0) = \frac{1}{2}$. In a more general framework note that this might occur whenever $q^*(0) > \frac{1}{2}$.

[28] An interesting consequence of this is that under weak coercion reliance without strong coercion reliance a gradual reduction of π toward 0 will lead to state failure, while a sudden change from π_0 to 0 will not.

bounds of $\mathbf{S}_{\pi_0}$ will be at the bounds of the $[0, \pi_{max}]$ interval. We define the different types of bounds on $\mathbf{S}_{\pi_0}$ as follows:

Definition of constraints on the set of implementable coercion levels: For any π_0, any upper or lower bounds of the set of implementable coercion levels from π_0, $\mathbf{S}_{\pi_0}$, will be either:

I. A *strategic* constraint; if the coercion level at the bound of $\mathbf{S}_{\pi_0}$ is a fix-point of the insurrection constraint; π_{fix}.

II. A *feasibility* constraint; if the coercion level at the bound of $\mathbf{S}_{\pi_0}$ is π_{max}.

If the upper bound of $\mathbf{S}_{\pi_0}$ is a strategic constraint, defined as $\overline{\pi_{fix}} \equiv sup\{\mathbf{S}_{\pi_0}\}$, it must lie at a crossing of the fix-point-line, by the composite insurrection function from above. An upper strategic constraint on $\mathbf{S}_{\pi_0}$, $\overline{\pi_{fix}}$, will be an attractor fix-point with $\hat{\rho}'(\pi) < 1$; the authority can increase π until it arrives at this level of coercion and will do so if this is the minimum of $\mathbf{Q}_{\pi_0}$. This property arises from that $\hat{\rho}(\pi)$ is continuously defined, hence any coercion level π' in the open $\mathbf{S}_{\pi_0}$ set must satisfy $\hat{\rho}(\pi') < \pi'$, while a strategic constraint by definition is on fix-point-line where $\hat{\rho}(\pi_{fix}) = \pi_{fix}$. The opposite holds for a strategic constraint at the lower bound of $\mathbf{S}_{\pi_0}$, defined as $\underline{\pi_{fix}} \equiv inf\{\mathbf{S}_{\pi_0}\}$, since the insurrection function has a lower bound that is a crossing of the fix-point-line from below, and it holds that $\hat{\rho}'(\underline{\pi_{fix}}) > 1$. Hence $\underline{\pi_{fix}}$ is a repeller fix-point; the response to use of coercion change the insurrection threshold sufficiently that the authority can increase π from $\underline{\pi_{fix}}$ to higher levels of coercion.

We now go on to find the optimal coercion level of the authority $\pi' \in \mathbf{S}_{\pi_0}$ by developing the notion of a dynamically stable equilibrium, π^{IC}:

Definition of a dynamically stable equilibrium: A dynamically stable equilibrium is defined as a coercion level and an SSE $\{\pi^{IC}, q^*(\pi^{IC})\}$ such that π^{IC} is the optimal coercion level if π^{IC} is equal to the initial condition π_0, $\pi_0 = \pi^{IC}$. [29]

This definition implies that a dynamically stable equilibrium, $\{\pi^{IC}, q^*(\pi^{IC})\}$, must fulfill the following three conditions; conditions 2. and 3. follow from 1. but are included for completeness:

1. π^{IC} is the solution to the static optimization of the authority (17) when $\pi_0 = \pi^{IC}$.

2. Socialization investment for both morality groups is unchanged at $q^*(\pi^{IC})$, i.e., $q^*(\pi^{IC})$ is an SSE.

3. The a morality group does not choose to commit an insurrection at π^{IC}; i.e., the coercion level does not breach the insurrection constraint; $\pi^{IC} \leq \rho(q^*(\pi^{IC}))$.

Conditions 2 and 3 in the definition of a dynamically stable equilibrium are fulfilled for all $\pi \in \mathbf{S}_{\pi_0}$, while condition 1 is fulfilled by the π in $\mathbf{S}_{\pi_0}$ that maximizes U^β. Hence, we find π^{IC} by solving:

$$\pi^{IC} \equiv \{\operatorname*{argmax}_{\pi} U^\beta = \min_{\pi \in \mathbf{S}_{\pi_0}} [\frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}]\} \tag{18}$$

---

[29]This is equilibrium in the infinitely repeated game where the authority first chooses π', whereupon a morality parents collectively choose whether to insurrect; if not, parents of both morality groups choose their levels of socialization investment τ^a, τ^b. New generations of parents then keep setting the socialization investment for each generation. When the population is in the new SSE, $q^t = q^*(\pi')$, the authority can set a new coercion level, and the game is repeated. Note that the parents only have preferences for their child's morality, not for any later generation, which simplifies the strategic aspects of the game.

Any authority that can infinitely reset π will always be at, or in a sequence $\{\pi_n\}_0^{N-1}$ leading up to, a dynamically stable equilibrium $\{\pi^{IC}, q^*(\pi^{IC})\}$. The dynamically stable equilibrium, π^{IC}, will be unique corresponding to every feasible π_0 except for one special case.[30] The monotone derivative of the insurrection function, $\rho'(q^*(\pi)) < 0$, implies that at any iteration there can never be another $q^*(\pi)$ giving a higher insurrection constraint than the lowest attainable $q^*(\pi)$. Hence, there can not be a lower reachable $q^*(\pi')$ than the π' reachable through minimizing $q^*(\pi)$ in every iteration. In other words, maximizing the capability to reach any long-term goal and maximizing short-term gains will imply equal behavior. The prediction of the dynamical equilibria is robust to the introduction of time preferences when coercion is costless and $\rho(q^*(\pi))$ is monotonically decreasing in $q^*(\pi)$.

Any dynamically stable equilibrium will either be at, a strategic constraint, a feasibility constraint or a local minimizer of $q^*(\pi)$. We establish this as Proposition 4:

Proposition 4. Dynamically stable equilibria For any initial condition, π_0, the dynamically stable equilibrium π^{IC} is a coercion level equal to either:

I. The unconfrontational level of coercion as an interior point of $\mathbf{S}_{\pi_0}$: $\pi^{IC} = \pi_{\underline{q}}$.

---

[30]The only instance in which the minimum of $\mathbf{Q}_{\pi_0}$ does not correspond to a unique minimum π' is the case that $\{\pi_{\underline{q}}, \pi_{\underline{q}}^e\} \in \mathbf{S}_{\pi_0}$, $\hat{\rho}(\pi_{\underline{q}}) = \pi_{\underline{q}}^e$ and $\pi_{\underline{q}}^e = \overline{\pi_{fix}}$; then by definition $inf(\mathbf{Q}_{\pi_0}) = \{q^*(\pi_{\underline{q}}), q^*(\pi_{\underline{q}}^e)\}$. Hence, $\pi^{IC} = \{\pi_{\underline{q}}, \pi_{\underline{q}}^e\}$ and the authority will be indifferent between imposing $\pi_{\underline{q}}$ or $\pi_{\underline{q}}^e$. In an application of the model this issue can be solved by considering whether the relevant authority has other considerations that make coercion costly, in which case $\pi^{IC} = \pi_{\underline{q}}$, or beneficial, in which case $\pi^{IC} = \pi_{\underline{q}}^e$.

II. A strategic constraint at the upper bound of $\mathbf{S}_{\pi_0}$: $\pi^{IC} = \overline{\pi_{fix}}$.

III. The upper feasibility constraint at the bound of $\mathbf{S}_\Pi$: $\pi^{IC} = \pi_{max}$. [31]

When $\{\Delta u, C(\pi)\}$ is sufficiently low such that $q^*(\pi)$ is of class III), π^{IC} will either be equal to a strategic constraint, $\overline{\pi_{fix}} \in [0, \pi_{\underline{q}})$, or, the lowest π' attaining $q^*(\pi) = 0$, depending on whether the composite insurrection function is such that π' is implementable from π_0. For $q^*(\pi)$ of class I), the dynamically stable π^{IC} is equal to $\pi_{\underline{q}}$ if this is implementable, and equal to some strategic constraint $\overline{\pi_{fix}} > \pi_{\underline{q}}$ if not. For $q^*(\pi)$ of class II), authorities will end up in stable gunpoint equilibria with two morality populations when $0 < q^*(\pi_{max})$ whenever π_{max} is implementable. If there exists an implementable π' such that $q^*(\pi') = 0$, then the population will go towards the single morality equilibrium. If this is not the case, then either π^{IC} is equal to $\pi_{\underline{q}}$, the unconfrontational level of coercion, or the equilibrium must be a strategic constraint, either at some coercion level above $\overline{\pi_{fix}} \in [0, \pi_{\underline{q}})$ or bellow, $\overline{\pi_{fix}} \in (\pi_{\underline{q}}^e, \pi_{max})$, the open interval of inefficient coercion levels, $(\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.

The equilibrium $\pi^{IC} = \pi_{\underline{q}}$ is the only one where the authority restraints its use of coercion in equilibrium; it imposes a coercion level strictly lower than the highest implementable coercion level. The equilibrium $\pi^{IC} = \overline{\pi_{fix}}$ is given by $\hat{\rho}(\pi)$ and im-

---

[31]Note that by definition of equilibrium as mutual reinforcing behavior *determining* an outcome, the equilibrium with the coercion level similar to the feasibility constraint at π_{max} could be omitted, since it is the feasibility constraint that determines the $\pi' = \pi^{IC}$, and it is not set at any threshold where π changes a strategic choice made by the a morality group to commit an insurrection.

41

plies a coercion level at a binding insurrection constraint. Finally; the equilibrium where $\pi^{IC} = \pi_{max}$ can be understood as an equivalent of legitimacy at the "barrel of a gun". The gun point level of legitimacy is defined as the legitimacy that can be achieved at the $q^*(\pi)$ corresponding socialization investment at its SSE value at π_{max}.

When a composite insurrection function is such that different initial conditions, π_0, generate different $\mathbf{S}_{\pi_0}$ sets, the model will have path dependency; different initial conditions will give different dynamically stable equilibria, π^{IC}. For a composite insurrection function, $\hat{\rho}(\pi)$, with two crossings of the fix-point-line such as the one described in Figure 1, $\mathbf{S}_\Pi$ will consist of two disjoint subsets, $\mathbf{s}$, one lower $\mathbf{s}'$ and one upper $\mathbf{s}''$; i.e., $sup\{\mathbf{s}'\} < inf\{\mathbf{s}''\}$. Whenever there is one or more unsustainable levels of coercion π' in between the upper, $\overline{\pi}$ and the lower $\underline{\pi}$ limits of $\mathbf{S}_{\pi_0}$; $\underline{\pi} < \pi' < \overline{\pi}$, there will be two or more sets of implementable coercion levels for any $\hat{\rho}(\pi)$. Assume that $\overline{\mathbf{S}}_{\overline{\pi_0}}(\pi) = \{\mathbf{s}'', \mathbf{s}'\}$, then $\underline{\pi_0} \in \mathbf{s}'$ and $\overline{\pi_0} \in \mathbf{s}''$, will produce different sets of implementable coercion levels depending on whether there is any way of implementing the minimal π' in the upper sub-set $\mathbf{s}''$ from the lower $\mathbf{s}'$; that is if $\rho(inf\{\mathbf{Q}_{\underline{\pi_0}}\}) < inf\{\mathbf{s}''\}$.

Generally, lower sub-sets of $\mathbf{S}_\Pi$, such as $\mathbf{s}'$, will always be contained in higher sets of implementable coercion levels, such as $\overline{\mathbf{S}}_{\overline{\pi_0}}$. This arises since an authority never instantaneously triggers an insurrection by lowering the coercion level; only when coercion levels lower than the initial condition are unsustainable can reductions in coercion lead an insurrection. Hence; the path dependency only goes from low to high levels of coercion; an authority can be restricted to impose a lower level of coercion than it would otherwise be able to impose, but cannot be restricted to impose higher levels of coercion than it otherwise could due to historical factors. In other

words, the lower bound of $\mathbf{S}_\Pi$ is the lower bound of any $\mathbf{S}_{\pi_0}$. We conclude our formal analysis by developing path-dependency as a proposition:

Proposition 5. Path dependency If and only if there exist initial conditions, $\overline{\pi}_0 \neq \underline{\pi}_0$, such that the set of implementable coercion levels from $\overline{\pi}_0$ or $\underline{\pi}_0$ differ, $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$, will different initial conditions $\pi_0 = \overline{\pi}_0$ and $\pi_0 = \underline{\pi}_0$ give different dynamically stable equilibria; $\overline{\pi}_{\overline{\pi}_0}^{IC} \neq \underline{\pi}_{\underline{\pi}_0}^{IC}$.

The intuition for why π^{IC} will *always* be different for different $\mathbf{S}_{\pi_0}$ is as follows: when the composite insurrection function is such that the sets of implementable coercion levels are different, it implies that some levels are sustainable and contained in one of the sets, but unimplementable from the highest implementable coercion level in the lower set. Since $\rho(\cdot)$ is homogenously increasing, this implies that one of the sets must contain a higher π corresponding to a lower $q^*(\pi)$ than the other; the dynamically stable equilibrium will always be the lowest attainable $q^*(\pi)$; π^{IC} is the minimum of $\mathbf{Q}_{\pi_0}$ and must consequently be different for the two different sets; $\overline{\mathbf{S}}_{\overline{\pi}_0}, \underline{\mathbf{S}}_{\underline{\pi}_0}$.[32]

---

[32]Note that we also developed that, equivalent to different sets of implementable coercion levels is that the lowest coercion level in the sub-set containing the coercion level equal to the upper dynamically stable equilibrium, $\overline{\pi}_{\overline{\pi}_0}^{IC}$, must not be implementable from the lowest implementable $q^*(\pi)$ from the other initial condition; $inf\{\mathbf{Q}_{\underline{\pi}_0}\}$. This $inf\{\mathbf{Q}_{\underline{\pi}_0}\}$ is by definition the $q^*(\pi)$ at the dynamically stable coercion level for this initial condition, $\underline{\pi}_0$. Defining the subset $\mathbf{s}'' \in \mathbf{S}_{\overline{\pi}_0}$ as the subset containing $\overline{\pi}_{\overline{\pi}_0}^{IC}$, we can develop the following statement of Proposition 5: Dynamically stable equilibria differ, $\overline{\pi}_{\overline{\pi}_0}^{IC} \neq \underline{\pi}_{\underline{\pi}_0}^{IC}$, if and only if there exist two initial conditions $\overline{\pi}_0 \neq \underline{\pi}_0$ such that $\hat{\rho}(\underline{\pi}_{\underline{\pi}_0}^{IC}) < inf\{\mathbf{s}''\}$.

Consequently, the model makes the prediction that coercion levels and corresponding SSE prevalence of morality groups will in some cases be inherently dependent on history in conjunction with the included long-term equilibrating factors.

<div align="center">❖</div>

The crucial assumption behind the developed results is that some levels of coercion are marginally effective and that at least one level of coercion in between these levels is strictly marginally ineffective.

We note that any constraint on coercion can be categorized as either a feasibility or a strategic constraint, in the sense that it will either be endogenously dependent on $q^*(\pi)$ or it will not. Further, note that any interval of coercion $[0, \pi_{max}]$ giving a class I) functional form of $q^*(\pi)$ can be seen as a feasibility constraint in some larger interval $[0, \pi_{max}]$ of class II) functional form.[33] We can sum up two main insights from the model in as follows:

Main insight on coercion use: A legitimacy-maximizing authority will restrain coercion of a morality when non-convexities in the response to coercion makes imposing more efficient coercion levels either strategically or feasibly unimplementable.

Main insight on path dependency: Dynamically stable equilibrium is inherently given by the history of the polity whenever there exist coercion levels that are

---

[33]Since coercion resentment is sinking in the last part of the interval $C''(\pi) < 0$ for $\pi \in (\hat{\pi}, \pi_{max}]$ any $q^*(\pi)$ of class I) must converge towards class II) as $\pi_{max} \to \infty$. This implies $\pi^{NC} = \pi_{max}$ as $\pi_{max} \to \infty$.

sustainable but unimplementable from some initial conditions. [34]

Further, we have established that strategic constraints must be fix-points of the composite insurrection function, established mechanisms for how polities can converge into state failure when coercion levels are unsustainable and established a formal definition of, and some necessary conditions for, weak and strong coercion reliance.

Most results should naturally generalize for other non-linear functional forms where marginal effectiveness of coercion varies with its level; i.e., dynamically stable equilibria will either be at a π_{fix} or at a local or global minimum of $q^*(\pi)$ within $[0, \pi_{max}]$. For instance; assuming an S-like coercion resentment function tantamount to the one assumed, the results of the analysis would naturally generalize onto this functional form; for each sufficiently concave interval of $q^*(\pi)$ there would be an additional inefficient interval of coercion and for each sufficiently convex interval that would be an additional internal stable equilibrium.

4. The theory applied to history

We now proceed to consider the religious homogenization in Early Modern Europe (1517-1685) and The Soviet Secularization project (1922-1991) through the lens of the model. Dynamics of changing moralities mainly take the perspective of rational elites seeking to maximize their influence, where the sentiments of the population are seen

---

[34]Note that, whenever multiple dynamically stable equilibria exist, an exogenous temporary increase in insurrection capability can move the dynamically stable equilibrium between sets of long-term implementable coercion levels.

as prerequisites to their strategies. The analysis assumes instrumental motivations rather than idealistic motivations for imposing the morality of the state upon the population. This is a simplifying assumption about intentions, which by nature will ultimately be difficult to prove or refute, as it is challenging to distinguish between how religious and ideological differences serve as motivation or justification for actions. Note that if one assumes elites to be fully intrinsically motivated, their motivation will often be to implement the authorities' preferred morality as an end in itself; which coincides fully with the proposed utility function of maximizing its prevalence. In other words, regardless of why the authorities wish to maximize their morality, the analysis remains the same as long as the authorities see it as a policy objective to maximize the prevalence of individuals that share the morality of the state.

4.1. Religious tolerance in early modern France and the Holy Roman Empire 1517-1685

The Christian Schism after the Reformation in 1517 and the subsequent spread of the Protestant faith, fueled by the introduction of the printing press (Rubin, 2014) and the dismay with policies of the Catholic Church, lead to an increase in religious heterogeneity in Early Modern Europe. The French kings and the Holy Roman Emperors (β) built their legitimacy on the Catholic Church (b) and the introduction of Protestantism (a) posed a threat to the legitimacy of their states (Johnson and Koyama, 2013).[35] In the first part of Early modernity (1517-1618), both the French and Holy Roman rulers saw a spread of Protestantism, combined with local nobility reforming to the Protestant faith to gain regional independence, leading to religious

---

[35]For the ease of presentation we do not distinguish between different Protestant faiths, i.e., Lutheranism and Calvinism.

46

civil wars. This new religious cleavage enhanced existing ongoing processes of regional fights for independence and removed the possibility of polyvalent religious legitimization; in France, against the backdrop of a growing absolutist state, in the Holy Roman Empire, against the backdrop of a fragmenting empire.

The initial religious wars and periods of upheaval ended with the admission of religious rights at the Peace of Augsburg (1555) in the Holy Roman Empire and the Edict of Nantes (1598) in France. These concessions were made as a response to the rulers' realization of the unproductiveness of the coercive measures $(C(\pi))$, coupled with an inability to sustain the ensuing military pressure (Wilson, 2009). As Johnson and Koyama (2013) put it: "This intensified persecution became increasingly ineffective: it served to strengthen the faith of Protestants and encouraged them to organize", pointing to the use of coercion as counterproductive; i.e., compatible with a micro level presence of coercion resentment.

Coercion resentment in the Holy Roman Empire

The Holy Roman Empire was not a unified state, but rather a decentralized empire structure of smaller kingdoms with a varying degree of loyalty to the ruling Habsburg family and the Holy Roman Empirical authorities. Protestantism served as both a cause of and an excuse for peripheral resistance against the central authorities; lower level princes actively used the religious cleavages and changed their religious affiliations to challenge the hegemony of the Emperor, build alliances and gain influence (North and Thomas, 1973). This demonstrates how religious homogeneity was a necessity for maintaining a strong state, and why implementing the morality of the state was seen as imperative for preserving the Empire united and under the control

of the ruling Habsburg elite.

After granting Protestants (*a* morality) the right to practice their faith at the peace of Augsburg; the Emperor Charles V (1516-1556) still saw Protestants as a challenge to his powers, and, at his death, the Habsburg family was divided between moderates and traditionalists in its view of which policies should be adopted towards the Protestants. The Habsburgs recognized the current coercion level was in the inefficient $(\pi_{\underline{q}}, \pi^e_{\underline{q}})$ interval, but were uncertain and divided on the direction in which to go, and whether coercion levels beyond $\pi^e_{\underline{q}}$ would trigger an insurrection.

The moderates wanted to pursue a non-confrontational line, building the legitimacy on both faiths, while the traditionalists saw the Habsburgs as having an intrinsic calling to be the champions of the Catholic faith in Europe and wished to purge the empire of Protestantism through the use of force, i.e., to heighten π.[36] In the language of the model, the Habsburg seemed to see their policy options as either a non-confrontational approach or strong and confrontational levels of coercion which they knew would be at the edge of, if not beyond, the set of their implementable coercion levels $\mathbf{S}_{\pi_0}$. History would prove that the Habsburgs did not have the internal military or the external strategic position to impose the coercion levels that they attempted.

---

[36]Emperor Rudolf II (1572-1608) conducted a conciliatory policy towards the Protestants seeing an alliance as a way to unify the Empire (Helfferich, 2009). He was, however, an introvert and an ineffective emperor and during his reign imperial influence deteriorated. Similarly, Emperor Mattihas' short reign (1612-1618) ran consolatory policies toward the Protestants.

In 1618, it became clear that Ferdinand II, who had pursued strong anti-Protestant policies in Austria, would be the successor to the throne; this further increased tension in the Protestant dominated region of Bohemia. In the period from the Peace in Ausburg in 1555 to 1618, the Protestants increased their numbers (Cantoni, 2015). This is compatible with an increased investment in socialization and consequently $q^*(\pi)$ in line with a micro-level coercion resentment. The imminent coronation greatly increased the resentment towards the empire and the anticipated change toward a more confrontational policy. Through the lens of the model, this can be seen as tipping the insurrection constraint following resentment towards the emperors' new-found ambition of a Counter-Reformation; it clearly acted as a prerequisite for the approaching conflict. The renewed program of confrontational religious homogenization that was anticipated after the coronation of the more religiously dedicated Emperor Ferdinand II, strongly contributed to the protestants' insurrection at the Second Defenestration of Prague sparking the Thirty Years' War (1618-1648) between Protestants and Catholics in Germany (Wilson, 2009).

Initially, Emperor Ferdinand II was successful in suppressing the Protestants; in the terminology of the model he had overstepped his insurrection constraint, but was able to stamp down the ensuing rebellion. The Empire won the first part of the war ending with the Battle of White Mountain (1620). Ferdinand II might have succeeded in uniting the Empire under one faith had it not been for the foreign involvement in the conflict. Foreign powers joined the conflict and sided with the Protestants to further their own causes; Protestant Denmark-Norway feared that a Catholic victory would threaten its sovereignty, Sweden feared that a strengthened emperor would ally with Catholic Poland to reclaim the Swedish crown, while the Catholic French aided the Protestant rebels in the Holy Roman Empire to weaken

the Habsburg Empire and deter the elite from supporting their Habsburg relatives reigning in Spain.[37] The Holy Roman Empire lost the war; Emperor Ferdinand II's violation of the insurrection constraint of the Holy Roman Empire in 1618 is, together with the Peace in Augsburg, responsible for Germany remaining a religiously divided country to this day.

Religious cleavages worked in furthering other strategic agendas for neighboring authorities as a legitimate way to engage in foreign conflicts (Nexon, 2009). These factors were to change with the new inter-state institutional order that was established as the war ended with the Peace of Westphalia, contributing to that while Ferdinand II's attempt to remove Protestantism ultimately failed, the French crown would succeed forty years later.

Religious persecution of the French Huguenots

Similarly to the wars of religion in the Holy Roman Empire, the French crown had waged war with its Protestant population, the Huguenots, from the beginning of the Reformation (1517). Recognizing the unproductiveness of its policies during the French Wars of Religion, the French crown settled for a non-confrontational equilibrium with the Edict of January at St.Germaine in 1562. Protestantism was decriminalized, but the Huguenots were not allowed to worship publicly; an illustrative example of a non-confrontational level of coercion, $\pi_{\underline{q}}$, in our model. The period following that of the study by Johnson and Koyama (2013), was generally a period

---

[37]Ringmar (2007) refutes this explanation of the Swedish rationale and argues that the Swedish elite embarked on the mission to Germany to be taken seriously as a Protestant European power. Nevertheless; the Swedes used a regional conflict to further their own agenda.

of increased prosecution of religious minorities following increased state capacity.

Prior to the decision to once more outlaw Protestantism with the Revocation of the Edict Of Nantes in 1686, advisors close to the French king, Louis XIV, recognized the potential counter-productiveness of this policy (Sutherland, 1988). Initially, measures were introduced gradually, as French historian Elisabeth Labrousse puts it: "measures, therefore, had to be constantly presented, albeit with a good deal of sophistry, not as aggressive sanctions but simply as a withdrawal of the kings' favors from the minority." (Sutherland, 1988). Marginal changes to the coercion level were gradually imposed to reduce salience and potential counter reactions. Louis XIV's advisors recommended a continuation of this policy by identification of the Huguenots as schismatic, a more gentle measure than outlawing Protestantism. However Louis XIV chose the stricter, more confrontational, line and the death penalty for Protestants was introduced in France on 1 July 1686. Consequently, it seems that the king's advisors did somehow recognize the potential for a reaction to his policies; the king was surprised, however, by the actual negative response and mass exodus; his hopes had been for reformation rather than relocation as a Protestant response.

While the granting of religious rights in 1547 to the Huguenots, was given in order to make peace with a politically and military powerful group, the revocation of the rights was made to a small group that posed little or no military threat following a long and gradual increase of coercive measures by the French Crown (Rae, 2002). Thus, in the terms of the model, since $\pi_{\underline{q}}$ was not the upper limit of the set of implementable coercion levels, the crown did not hit any fixed points of the composite insurrection function along the program of increases in π towards the dynamically stable equilibrium, which would turn out to be the gunpoint equilibrium. In line

with Proposition 2, once the coercion levels approached an inefficient level, Louis XIV went directly to a clear case of a gun-point threat avoiding any potentially inefficient coercion levels in $(\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.

Scholars studying the period surmise that without the actions of the state, the Huguenot identity might have withered away in the absence of persecution (Labrousse, 1985 in Rae, 2002). The identification of the emigrated French Protestants as Huguenots; a separate identity from the Catholic French, would remain strong, albeit outside France (Sutherland, 1988). This insight is interesting in the light of the model; it points to either the "cultural memory" of persecution, $C(\pi)$, having a long-term identity building effect after its immediate effect, or a persistence in high investment in socialization. The policy did not come at any military cost to the French king but had a reputational cost. Reactions from foreign kings were negative, condemning the treatment of the French Huguenots (Labrousse, 1985 in Rae, 2002), perhaps pointing to nascent expectations of minority rights being respected in international relations.

A comparative perspective on religious persecution in the Holy Roman Empire and France: The role of the Peace of Westphalia

The aftermath of the Thirty Years' War, the Peace of Westphalia, is thoroughly studied in international relations and considered to be the start of the modern state system. Interpreting the new institutional paradigm of international relations in Europe as relaxing the insurrection constraint, $\rho(q^*(\pi))$, due to a lower risk of foreign involvement, explains the freezing pattern in the map of religious identities after 1648 in Europe, thoroughly documented in political science (Rae, 2002; Nexon, 2009; Tilly

and Ardant, 1975; Rokkan, 1999).

The human suffering of the war, which killed an estimated third of the population within the borders of modern-day Germany, increased both the demand for a new paradigm and the respect for the new institutional rules of international relations in Europe.[38] Further, at an elite level, The Thirty Years War was seen as how not to wage war; an example of the dangers of religious passions and mercenary armies (Philpott, 2001).

Amongst the changes agreed at Westphalia was the principle of territoriality which created at least a minimal requirement of a legitimate claim to territory. It tied the religious identities to territorial identities, increasing the need for religious homogeneity. The treaty obliged the king to have the same religious affiliation as that of his polity (Wilson, 2009); thus reducing the incentives of changing faith to gain power. Further, it outlawed the use of religious tension in neighboring countries as a legitimate reason for engagement in civil wars. Together with the further delegitimization of mercenary armies, these measures effectively decreased the insurrection constraint, as religious minorities lost their role as a potential for furthering the interest for foreign powers as strategic "Jus ad bellum"; a pretext to go to war.

The model points to how lower insurrection constraints will lead to lower prevalence of non-state morality either through a quicker convergence towards an equilibrium or by enabling the authority to impose a program towards the gunpoint equilibrium. Hence, the model can account for how the Peace of Westphalia increased internal

---

[38]Estimates range from 10 percent to 45 percent (Theibault, 1997).

homogenization as a consequence of the delegitimization of using religious schisms as a pretext for foreign involvement in internal conflicts.[39]

While the attempts to homogenize the Holy Roman Empire lead to insurrection, foreign involvement and subsequently religious division of the Empire, the potential Huguenot mobilization could not be turned into a pretext for foreign involvement and a consequent military threat to the French king under the new institutional framework. This absence of threat from neighboring countries greatly relaxed the insurrection constraints as governments could focus on internal enemies when pursuing homogenization, thus predicting closer alignment between territory and state moralities (Nexon, 2009).

The changing military technology, away from professionalized soldiers with training in the use of both firearms and swords, towards mass armies primarily reliant on gunpowder, placed a higher military value on draftable citizens.[40] In the language of the model, changing military technology lead to a higher $\rho'(q^*(\pi))$; the insurrection constraint became more sensitive to the prevalence of the state morality as military capabilities became more sensitive to mass support. This, in turn, lead to an increase in the demand for homogenization of populations, enabling the draft large standing

---

[39]The potentiality of foreign powers using religious schism to legitimize military action in Continental Europe, and the absence of this risk in Britain as it is an island nation with strong natural borders, might potentially provide another piece of the puzzle of understanding the comparatively early emergence of elite intentions to achieve religious tolerance in Britain.

[40]The empirical relationship between military technology and the need for mass armies is discussed widely, from the classic Roberts (1954); to recent economic literature, see for instance Onorato, Scheve and Stasavage (2014).

armies against external threats, which would propel the development of consolidated states.

4.2. The Soviet Secularization project 1922-1991

The Soviet Authorities (β) had a clear and stated agenda to reduce the prevalence of religious morality (a), and used coercion (π) against the major religions of the USSR, Christianity and Islam, in order to increase the prevalence of its own secular morality, communism (b).[41] Similar to that of the Catholic kings of the Early Modernity, the approach towards religious communities in the USSR was initially very oppressive. The Great Terror of the 1930's saw widespread killings and forced Gulag encampment of religious individuals failing to denounce their religion. From 1937 onwards the Soviet Authorities altered their policies towards religion. The combination of the strong coercive measures proving inefficient according to the Soviet authorities' own 1937 consensus, and the need to apply religious and national sentiments at the beginning of WW2, moved secularization measures from severe and strongly coercive, to unconfrontational and less malignant (Froese, 2008).[42]

Data adopted from Froese (2008) shows how religious morality (q) in USSR decreased as a consequence of the deliberate Soviet policies to reduce its prevalence

---

[41]Implicitly we here assume that Communism can be understood as a set of internalized values on par with religion; indeed the Soviet authorities themselves saw it this way Kula (2005).

[42]Illustrative of the approach of the authorities are the names of the atheist movement founded by the USSR authorities; before 1920 the organization of atheist was named League of Militant Atheist, literally translated from Russian, League of the Militant Godless, which was disbanded at the onset of WW2 when the secularization project was put on hold. The later atheist organization founded after WW2, was named the Knowledge Society.

55

(see Figure 3.), while it increased again after the fall of the Soviet Union following the cessation of anti-religious policies (π) (see Figure 4.). All in all, the attempt to secularize the Christian regions of Soviet society was successful in that it lead to a drastic reduction of the prevalence of religious morality, but it did not lead to a full removal of religious sentiment.

The Soviet authorities and Christian churches

From 1937 onwards the major Christian churches of the Soviet Union were able to continue their practice, albeit facing censorship and demands from the authorities to serve the purposes of the Communist Party. The high degree of organization and internal hierarchies meant that both the Russian Orthodox Church and the Catholic Church were forced to become integrated into the Soviet system and continue their practice facing strong censorship. Protestantism and other less hierarchial Christian communities were often strengthened by the feeling of spite towards the Soviet authorities' anti religious policies, i.e., a response to the level of coercion. To the extent that religion persisted in the predominantly Christian parts of the Soviet Union, it did so largely by the use of what Greif and Tadelis (2010) refer to as crypto-morality; hidden from the public eye.

The persistence of religion was stronger in the areas where churches were aligned with other cleavages relative to the Russian amalgamate of identities associated with the Soviet rule. This was especially true where the church was seen as opposing the state; one example is the membership to the state autonomous Lithuanian Roman
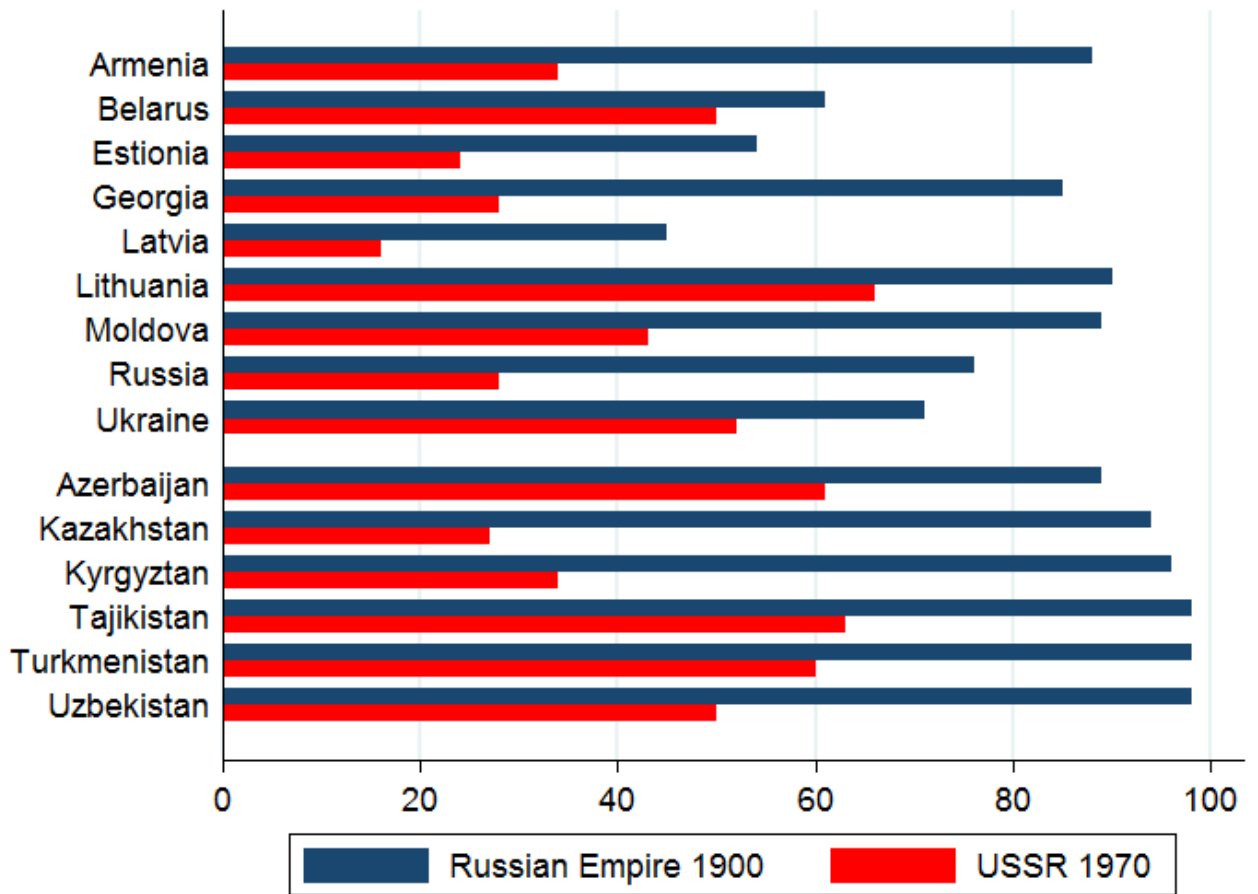
Figure 3: Religious affiliation with all religions before the Soviet Union, during the Russian Empire (1900) is shown as the bottom blue bars and under during the Soviet Union (1970) is shown as the top red bars
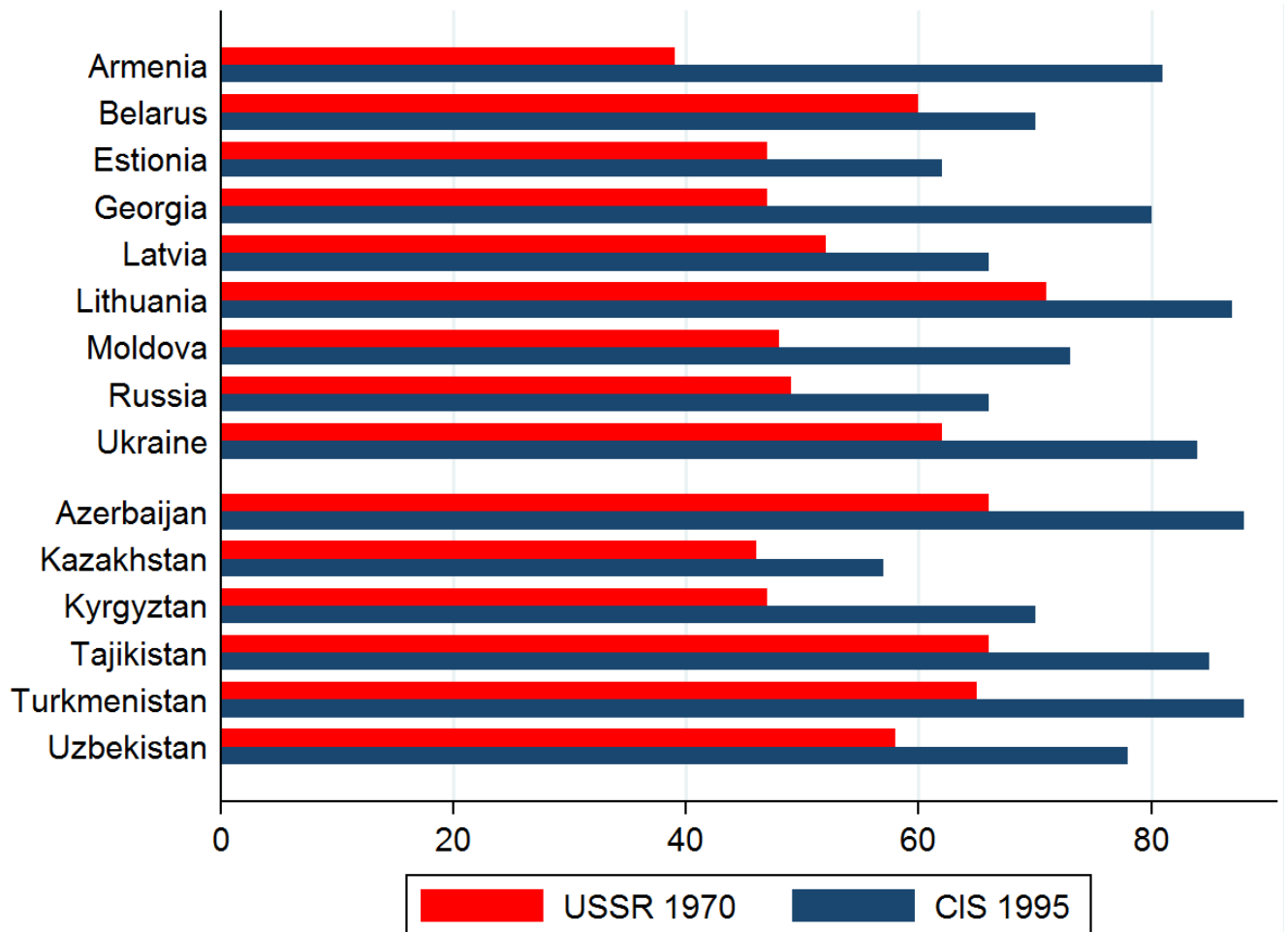
Figure 4: Affiliation with the majority religion under the Soviet Union (1970) is shown as the top red bars and after the Soviet Union under the Commonwealth of Independent States (CIS) (1995) is shown as the bottom blue bars. Source: Froese (2008). All Christian Countries have Orthodox Christianity as majority religion except Lithuania which is Roman Catholic and Latvia which is Lutheran, while all Muslim countries in the Soviet Union are of Sunni Islam as their majority religion except for Azerbaijan, which is of the Shia Islamic faith.

Catholic Church which was seen as synonymous with resistance to the Soviet authorities Froese (2008). This suggests to that where the framing of religious persistence was aligned with other in-group, out-group dynamics, Δu towards the secular USSR identity was higher. A possible explanation is that, as the framework predicts, a higher Δu lead to a higher q_t, which gave members of these communities a higher utility of rejecting the authority in terms of social recognition and lead to more visible resentment towards anti-religious policies.

Coercion resentment in Central Asian Soviet

The secularization policies towards the Muslims in Central Asia were even more cautious than in the predominantly Christian Orthodox part of the Soviet Union. In Central-Asia, insurrection risk was higher due to the weaker military presence and larger cultural differences towards the secular Communist morality, Δu. The potential gains for the local population arising from the Soviet rule, from modernization and economic development, was higher than in the Baltic regions. Hence, in accordance with predictions from the classic Nash bargaining model (Nash, 1953), Communist and local leaders on both sides had poor outside options and better incentives to cooperate; contributing to the comparatively more benign elite climate of communication than that between Moscow and the Baltic elites.

Froese (2008) describes how Soviet and Muslim authorities found common ground; although the Communist agenda in the long run was to destroy Islam, which they saw as prejudice against reason, Lenin described "Muslim folk heroes as emblematic of the human struggle against oppression", while Muslim Scholars noted that Islam could justify "even the rule of an usurper as means of assuring the public order and

59

the unity of all Muslims". The tone between the Soviet and the Muslim Authorities can be read between the lines in a letter from the Central Religious Muslim Board in 1942 to Stalin; "...champion of liberation of oppressed peoples and a man ever attentive to the need of the peoples...May Allah bring your work to a victorious end."(Marshall, Bird and Blane, 1971). Implicitly, the council signals that they are sympathetic to Stalin's cause, but that he will not succeed without the assistance of Allah (Froese, 2008).

The Soviet Authorities framed Communism as modernization, sweetening the deal of Soviet rule with promises of economic development to gain the partial support of Islamic intellectuals in Central Asia (Northrop, 2001). Policies such as the removal of Muslim courts were cautiously framed as modernization and done in cooperation with moderate Islamists. Stalin initially allied with Muslim modernization movements, most notably the Jadidism movement that sought to "rationalize Islam, to purify it and bring it into line with the modern era" through "progress, development and growth". Although the secularization of The Central Asian USSR was deliberately non-confrontational, there were, however, clear reactions to the Soviet anti-religious policy. An illustrative example of this is the violent reactions to the 1920's Hujum policy of having Muslim women remove their veils (Northrop, 2001).

Stalin would later deceive his former Jadid allies and purge most of its leadership on suspicion of their ambitions for further national independence for the Central Asian republics (Bennigsen and Lemercier-Quelquejay, 1967). Stalin's fears of growing demands for autonomy were not unfounded; Bennigsen and Lemercier-Quelquejay (1967) describe how Muslim national identities that were barely present in 1917 emerged in part as a result of the anti-religious policies to gain increasing salience

60

in 1967. They account for this effect by what they describe as "resentment against cultural and administrative domination of the Russians"; sentiments that could be turned into momentum for an insurrection against the USSR.

However, the promises of development and growth were not reneged by The Soviet Authorities; they trusted that their Muslim counterparts would not attempt to secede enough to endow them with a more working economy. This growth happened alongside positive social changes in Central Asian USSR, for instance women were given a comparatively independent role and educational levels increased, further integrating Central Asian USSR with Moscow. Together with the strengthening of the military capability of the Russians, these changes made any threat of cessation less realistic (Conquest, 1970).

Stalin either persecuted in a heavy-handed manner or kept a non-confrontational approach. This implies that he avoided a mid interval of coercion, in line with the model's prediction of authorities avoiding the inefficient $(\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ interval where latent strategic constraints are in place. While the treatment of Muslims in Central Asia was relatively benign, the treatment of smaller groups of Muslims in the southwestern region of Russia, such as the Crimean Tartar and the Chechens, was much more coercive and confrontational; forced deportations and subsequent expropriation of land were the primary instruments (Conquest, 1970). The potential threat of the Russian Muslim population in the south-west, and Caucasus allying with the Germans, were used as a pretext for the deportations; but this motive cannot explain that the differential treatment persisted after the end of WW2. The comparative differences may be explained by the Soviet authorities being aware of an inefficient interval of coercion $(\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ and restraining their use of coercion as a response to an

61

insurrection constraint in Central Asia, while pursuing levels beyond $\pi_{\underline{q}}^e$ in Europe, having no strategic constraints here. Assuming that Stalin perceived the response to coercion as stable across regions, this historical evidence supports Proposition 2 and indicates that the combination of cultural differences and negative response to coercion was sufficiently hostile as falling in under class II) in the USSR.[43]

5. Concluding remarks

We have developed a model that demonstrates how the micro-foundation of coercion resentment can be used to understand the macro dynamics of legitimacy maximizing authorities. We have argued that the assessed monarchs of the Early Modernity and Stalin restrained their use of coercion in response to strategic constraints in a way that is explainable by our theoretical framework.

The model implicitly assumes atomized agents and abstracts from dynamics of legitimacy caused by communities, organizational structure, framing or strategic interaction between elites. These implicit simplifications are justified as long as community leaders are equally good at maximizing their own influence through playing on salient cleavages. If both moralities have leaders that frame situations equally well in terms of creating saliency, then the underlying potential for a cleavage will be the relevant mechanism at play. In other words; if one pictures the "facts on the ground", i.e., the actual given potential for action, the cards in the hands of the community leaders,

---

[43]The disparately harder treatment of the Eastern Muslims continued after the end of WW2, pointing to an additional consideration for Stalin's differential treatment, the heightened need for high legitimacy and stronger capability to coerce in Europe. Stalin also had an additional value from coercing the Eastern Muslims, as it sent a credible signal about the cost of collaboration with competing authorities to other minority communities.

then, if, on average, they play their cards equally well, the mechanisms in the model will be the driving factors.

There are several potential extensions of this model that can address related questions in future research:

Parental preferences for coercion: An applied problem is considering that the state morality parents can fully or partially set the coercion level. Parents of the state morality will need lower levels of parental socialization to attain their preferred morality in an equilibrium with positive levels of coercion.[44] Assuming that the majority of parents do not have utility over the outcomes of the state, parents of the state morality will prefer the coercion level that balances the tradeoff between the private preferences of lowered socialization and the social preferences for future generations of state and non-state morality children and parents. Exploring a model where parents can set coercion levels in conjunction with historical evidence can shed light on processes where democracies become coercive or authoritarian. Further theoretical work along these lines can address the question of to which extent totalitarian policies emerge from political demand or political supply.

Evolutionary properties of state competition: Note how the model predictions hold in a framework where authorities are naïve about the effect of coercion; authorities that impose coercion levels within the set of implementable coercion levels will endure, and others perish from insurrections. Further theoretical inquiries that apply the

---

[44]This holds as long as the state morality is also the majority morality. If the state morality is a minority the issue depends on functional form.

set of implementable coercion levels can tie together empirical evidence of historic and pre-historic processes of state competition in new ways. For instance, assume an extension of this model where populations of polities of uniform size and initial condition compete. Assume that the authorities are naïve about the effect of coercion but able to use military capabilities externally to overtake neighboring states. The population of polities in such a model will, presumably, over time converge towards only polities imposing the dynamically stable equilibria; room for deviation from the optimal polices will grow with differences in the relative sizes of polities and their initial conditions. Hence, it appears that the proposed equilibria can arise from state competition, in line with the arguments set forth in Tilly (1992); even under naïvité of the effects of, and constraints on, coercion.

Costly coercion under discounting: Exogenous variations in insurrection costs, variations in benefits of legitimacy of the authority and varying benefits of seceding for the minority can arise from factors such as rough terrain or rents from natural resources. Hence, there are reasons to assume that the set of implementable coercion levels might be different for authorities with availability of equal military technology, and that authorities might choose to impose different coercion levels due to differences in benefits, or costs, of legitimacy.[45] Further, under costly coercion, equilibrium outcomes will also be given by the authority's time preferences of the authority; there

---

[45]An extended model that includes these properties could provide a micro foundation to Barfield (2010)'s explanation of the high ethnolinguistic variance in Afghanistan. He places emphasis on how rough terrain, giving a low insurgency cost compared to the low value of attaining legitimacy, together with multiple historic influences, i.e., multiple seed moralities, and low benefit of having legitimacy, have contributed to the large cultural heterogeneity observed in Afghanistan.

will be a trade-off between the discounted future benefit of legitimacy and present cost of coercion. This could account for why different dynastic structures, i.e., more or less direct hereditariness of power, could lead to different policies. In modern democracies such differences in incorporating the future can arise from variations between candidate politics versus party politics.

Strategic aspects of multiple competing authorities: Technology and the composition of multiple ethnic groups might make the set of sustainable coercion levels empty for any single authority; hence in some instances, creating strategic coalitions between authorities is a necessity for establishing a sustainable state. The model could be extended towards a cooperative game theory framework to analyze the strategic dimensions of internal and external competing authorities under varying insurrection constraints. This can address questions such as sustainable polity borders and how intervention in polities with multiple authorities should optimally apply the local power structures.[46] Further, the endogenous treatment of morality prevalence makes the framework able to identify when a peace agreement between competing authorities will not be long term sustainable, i.e., when long term changes in prevalence of moralities will affect power balances to render a previous agreement an out of equilibrium outcome.

---

[46]Expert on the state development in Africa, Robert H. Bates, predicts that the key to understanding the structure of wars in Africa versus Europe lies in understanding the Peace of Westphalia constrained the European elites in terms of using neighboring ethnic cleavages to further their cause (Weingast and Wittman, 2008). Use of ethnic conflicts in search for influence has generally been a major cause of instability in Central-Africa; one example being the conflict in Eastern-Congo (1998-2003).

Framing and timing of coercion: It seems likely that effects such as cultural memory, incentives of community leaders and sluggishness in investment in military technology, change the effect of coercion and consequently the set of implementable coercion levels. Further, different programs in terms of how gradual changes are and how they can be framed, will imply that the set of reachable coercion levels will be different for different strategies and different pre-histories. Explicit modeling of the effects of timing and framing of coercion can be applied to understand how short-term processes determine convergence towards the long-run equilibria.

Strategic conditions of coercion reliance: Which strategic pre-histories are conducive for authorities establishing coercion reliant states? Addressing this question can complement the rich and established literature on the path to inclusive institutions from a new angle; how did authorities strongly reliant on coercion arrive in this situation? Further it can potentially give theoretical insights on which paths of state development lead to malignant outcomes, and at what critical junctures these paths can be avoided.

The frameworks' explicit modeling of population responses' together with the possibility of strategic analysis, makes the framework a potential tool for policy analysis for an external agency constraining an authority's use of coercion. Generally, limiting the level of accepted coercion will depend on how external agencies consider the ratio of cost of commission versus the cost of omission, i.e., the cost of limiting coercion, and the benefit of limiting the suffering caused by coercion itself. Further research can form a theory that takes ethnic compositions and power relations as inputs to produce predictions on which initial states, i.e., polity borders that can create sustainable uncoercive states, what the cost of reaching these states is and

where the pitfalls of state failure lie.

Building and empirically investigating portable general models of these dynamics plays an important role in using the conflicts of the past to avoid conflict in the future, and to further understand how diversity of moralities can be an equilibrium outcome in the face of legitimacy-maximizing authorities. Although technology, beliefs, and institutions change, as long as human nature remains stable, the past will be informative of the future. From understanding democratic transitions to policy recommendations in states like Syria and Afghanistan at the time of writing; legitimacy and its dynamics remain an important phenomenon.

6. Acknowledgments

7. Appendix 1: Proofs

7.1. Proof of Lemma 1

Lemma 1: The smallest group always invests more in parental socialization; $\tau_t^b \leq \tau_t^a$ if and only if $q_t \leq (1 - q_t)$.

Proof: Suppose $q_t < (1 - q_t)$, it then follows from (6) that $H'(\tau_t^b) < H'(\tau_t^a)$. By the Inada condition of $H''(\tau^m) \geq 0$ in Assumption 3, it follows that $\tau_t^b < \tau_t^a$. The only if part follows from that there are only two moralities.

□

7.2. Proof of Lemma 2

Lemma 2: There is a unique stable interior SSE at $q^* = \frac{1}{2}$

Proof: We show that (i) there exists a unique interior SSE at $q^* = \frac{1}{2}$, (ii) and that it is stable.

(i) Existence and uniqueness of an interior $q^* = \frac{1}{2}$

A steady state equilibrium (SSE) level of q, denoted q^*, is reached when $q_t = q_{t+1}$. It follows from (3), $q_{t+1} = q_t + q_t(1 - q_t)(\tau^a - \tau^b)$, that for $q_t = q_{t+1}$ to be fulfilled, $q_t(1 - q_t)(\tau^a - \tau^b) = 0$ must hold. Hence at any interior SSE, i.e., $q^* \in (0, 1)$, $\tau^a = \tau^b$. From (6), it follows that this implies $q_t = (1 - q_t)$ which gives $q^* = \frac{1}{2}$.

(ii) Stability of $q^* = \frac{1}{2}$

We will show that for any $q_t \in (0, 1) \neq q^*$ there will be convergence towards q^*. Suppose $q_t > q^*$, it then follows from (6) that $H'(\tau_t^b) < H'(\tau_t^a)$. By Lemma 1 it follows that $\tau_t^b > \tau_t^a$. By (3), $q_{t+1} < q_t$ when $\tau_t^b < \tau_t^a$ and $q_{t+1} < q_t$ for $\tau_t^b > \tau_t^a$. Thus

any $q_t \in (0,1)$ will converge to q^*. In other words, $(0,1)$ is q^* basin of attraction.

$\square$

7.3. Proof of Lemma 3

Lemma 3: For all pairs of $\{\pi, \Delta u\}$ two exterior SSEs exist. For some, but not all, pairs of $\{\pi, \Delta u\}$ a unique stable interior SSE exists, given by $q^*(\pi) = \frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}$.

Proof: (i) For all pairs of $\{\pi, \Delta u\}$ two exterior SSEs exist.

By definition a SSE is given by $q_t = q_{t+1}$. For $q_t \in \{0,1\}$ (3) implies that $q_t = q_{t+1}$ for any pair of $\{\pi, \Delta u\}$.

(ii) For some, but not all, pairs of $\{\pi, \Delta u\}$ a unique stable interior SSE exists, given by $q^*(\pi) = \frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}$.

Suppose $\frac{\Delta u' - \pi' + C(\pi')}{2\Delta u' + C(\pi')} = \frac{1}{2}$ and that $\{\pi', \Delta u'\}$ is the imposed π and Δu. We now show that this implies there exists an SSE where $q^*(\pi') = \frac{1}{2}$.

Consider $q_t = \frac{1}{2}$. Since $\Delta u' - \pi' + C(\pi') = 1 > 0$, (10) implies that $\tau_a > 0$ for $q_t > 0$. Since $\Delta u' + \pi' > 0$, we see from (11) that $\tau_b > 0$ for $(1 - q_t) > 0$. For $q_t = \frac{1}{2}$ to be an SSE it follows from (3) that it must hold that $\tau_a = \tau_b$. This implies the left side of (10) should equal the left side of (11). Under $q_t = \frac{1}{2}$, this gives $\frac{1}{2}(\Delta u - \pi' + C(\pi')) = \frac{1}{2}(\Delta u + \pi')$. This implies $2\pi' = C(\pi')$, which is consistent with $\frac{\Delta u' - \pi' + C(\pi')}{2\Delta u' + C(\pi')} = \frac{1}{2}$.[47]

---

[47]This argument holds mutatis mutandis for any $q^*(\pi'') = \frac{\Delta u'' - \pi'' + C(\pi'')}{2\Delta u'' + C(\pi'')} = \frac{m}{n} \in (0,1)$ and $q_t = \frac{m}{n}$. Hence, for any $\{\Delta u'', \pi''\}$ such that $q^*(\pi'') = \frac{\Delta u'' - \pi'' + C(\pi'')}{2\Delta u'' + C(\pi'')} \in (0,1)$ an internal SSE exist.

Uniqueness of the interior SSE $q^*(\pi)$ is trivially given by that $q^*(\pi) = \frac{\Delta u - \pi + C(\pi)}{2\Delta u + C(\pi)}$ is a single-valued function. The equilibrium is stable as $(0, 1)$ is $q^*(\pi)$ basin of attraction following the lines of the argument in the proof of Lemma 2 part *(ii)*.

We finally show that for some $\{\pi'', \Delta u''\}$ no interior SSE exits. Suppose $\{\pi'', \Delta u''\}$ is such that $\Delta u'' - \pi'' + C(\pi'') \leq 0$. By (10) and the Inada assumption that $H'(0) = 0$ and $lim_{\tau \to 1} H'(\tau) = \infty$ it follows that $\tau_a = 0$ for all q_t. Since $\Delta u > 0$ by Assumption 1 it follows from (11) that $\tau_b > 0$ for all q_t. It follows from (3) that if $\tau_a \neq \tau_b$ for all q_t, no interior SSE exists.

□

7.4. Proof of Lemma 4

Lemma 4: Imposing a coercion level π' corresponding to an interior SSE, $q^*(\pi') \in (0, 1)$, from an initial interior SSE $q^*(\pi_0)$ will make q converge to $q^*(\pi')$.

Proof: Assume the population is in some interior $q^*(\pi_0)$ and at time $t = 0$ a $\pi' \neq \pi_0 : q^*(\pi') \in (0, 1)$ is imposed. Since $\pi' \neq \pi_0$ and there is a unique interior SSE by Lemma 3, the FOC conditions for an SSE cannot be fulfilled, i.e., $q_0(\Delta u - \pi' + C(\pi')) \neq (1 - q_0)(\Delta u + \pi')$, at time $t = 0$. This implies $H'(\tau_t^b) \neq H'(\tau_t^a)$ due to the Inada conditions on $H(\cdot)$ it follows that $\tau_t^b \neq \tau_t^a$, and by (3) it follows that $q_1 \neq q^*(\pi_0)$. We define the following sequence of $q_0, q_1...q_N$ values under π' as $\{q_0, q_1......q_N, \pi'\} \equiv \{q_t\}_{\pi'}$.

We first establish that *(i)* any q_t in $\{q_t\}_{\pi'}$ will move in the direction of $q^*(\pi')$, *(ii)* that no $q_t \in \{q_t\}_{\pi'}$ is equal to the absorbing state exterior SSE $q_t \in \{0, 1\}$, and finally *(iii)* that $q_t \to q^*(\pi')$.

(i) Any $q_t \in \{q_t\}_{\pi'}$ will move in the direction of $q^*(\pi')$

By q_t moving in the direction of $q^*(\pi')$ we mean that if $q_t > q^*(\pi')$ then $q_t > q_{t+1}$ and if $q_t < q^*(\pi')$ then $q_t < q_{t+1}$.

First, note that at $q_t = q^*(\pi')$, as established in the Proof of Lemma 3, it holds that $(1 - q^*(\pi'))(\Delta u - \pi' + C(\pi')) = q^*(\pi')(\Delta u + \pi')$ as follows from the proof of Lemma 3. Suppose $q_t > q^*(\pi')$. It then follows that $(1 - q_t)(\Delta u - \pi' + C(\pi')) < q_t(\Delta u + \pi')$ which by (10) and (11) implies $H'(\tau_a) < H'(\tau_b)$. It follows from the Inada condition of $H'(\cdot) > 0$ that this implies $\tau_t^b > \tau_t^a$. Suppose $q_t < q^*(\pi')$, then the opposite holds. By (3) it holds that $q_t < q_{t+1}$ when $\tau_t^b < \tau_t^a$ and $q_t > q_{t+1}$ if $\tau_t^b > \tau_t^a$.

(ii) No $q_t \in \{q_t\}_{\pi'}$ is equal to the absorbing state exterior SSE; $q_t \in \{0, 1\}$

We first show that an interior $q^*(\pi')$ implies positive levels of socialization for both groups at all $q_t \in \{q_t\}_{\pi'}$, then, we demonstrate that this implies no exterior $q_t \in \{0, 1\}$ is in $\{q_t\}_{\pi'}$.

Since $q^*(\pi') = \frac{\Delta u' - \pi' + C(\pi')}{2\Delta u' + C(\pi')} \in (0, 1)$ it must hold that $\Delta u' + \pi' > 0$ and $\Delta u' - \pi' + C(\pi') > 0$. The FOC conditions, (10) and (11) and the Inada condition $H'(0) = 0$, implies that $\tau_a > 0, \tau_b > 0$ for any $q_t > 0$ when $\Delta u' + \pi' > 0$ and $\Delta u' - \pi' + C(\pi') > 0$. Hence, there will always be $\tau_t^a > 0, \tau_t^b > 0$ under π' for all $q_{t-1} > 0$. Since $q_0 \in (0, 1)$ it follows that $\tau_a > 0 \ \tau_b > 0$ and $q_{t-1} > 0$ for all $q_t \in \{q_t\}_{\pi'}$.

From (3), $q_{t+1} = q_t + q_t(1 - q_t)(\tau^a - \tau^b)$, we see that an exterior $q^* = 0$ or $q^* = 1$ cannot be reached from any interior $q_t \in (0, 1)$ if $\tau^a > 0, \tau^b > 0$.

(iii) $q_t \rightarrow q^*(\pi')$

This proof applies Proposition 1 , Proposition 2 and the definition of cultural substitution in Bisin and Verdier (2001 p 303-307). Following the proof of Proposition 2 in Bisin and Verdier (2001), we show that socialization level, τ, as a function of q_t satisfies the definition of cultural substitution in Bisin and Verdier (2001). It then follows from Proposition 1 in Bisin and Verdier (2001) that this implies that $q_t \rightarrow q^*(\pi')$.

We define $q_t^a \equiv q_t$ and $q_t^b \equiv 1 - q_t$ and denote a portion of a morality m, q^m. The requirements for cultural substitution on page 303 in Bisin and Verdier (2001) can be stated that *(i)* $\tau^m = d^m(q_t^m)$ where $d^m(q_t^m)$ is a continuous function, *(ii)* $d^m(1) = 0$ and *(iii)* $d^m(q_t^m)$ is strictly decreasing in q_t^m.

(i) $\tau^m = d^m(q_t^m)$ is a continuous function.

From (10) and (11) it follows that:

$$\tau_a = H'^{-1}((1 - q_t)(\Delta u - \pi' + C(\pi'))) \tag{19}$$

$$\tau_b = H'^{-1}(q_t(\Delta u + \pi')) \tag{20}$$

We first show that $H'^{-1}(\cdot)$ is defined. First, note that the Inada conditions $H'(\tau) \geq 0$, $H''(\tau) > 0$ and $H'(0) = 0$, implies that $H'(\cdot) > 0$ for all other τ than $\tau = 0$. The Inada condition of $lim_{\tau \rightarrow 1} H(\tau) = \infty$ imply $H'(\tau)$ maps from $[0, 1) \rightarrow [0, \infty)$, $H'(0) = 0$ and $H''(\cdot) > 0$ implies $H'(\cdot)$ has a continuous positive derivative. Hence, for every q_t^m $H'(\cdot)$ assigns a unique value i.e., $H'(\cdot)$ is a one-to-one function with a defined continuous inverse, $H'^{-1}(\tau^m)$, mapping from $[0, 1) \rightarrow [0, \infty)$.

Since everything inside $H'^{-1}(\cdot)$ in (19), (20) but q_t remains fixed for all $q_t \in \{q_t\}_{\pi'}$ and since $q^*(\pi') > 0$ implies that $\Delta u - \pi' + C(\pi') = K^a > 0$, $\Delta u + \pi' = K^b > 0$ we

can define $H'^{-1}(q^m K^m) \equiv d^m(q^m)$. Since $q_t \in (0,1)$ we can write $\tau^a = d^a(q_t)$ and $\tau^b = d^b(1 - q_t)$.

(ii) $d^m(1) = 0$.

Following from (8) and (9) parents are indifferent between choosing some infinites-imal amount of socialization and no socialization for $q_t \in \{0,1\}$. By assumption in footnote 9 on page 18 we have assumed that it holds that $\tau^a = 0$ for $q_t = 1$ and $\tau^b = 0$ for $(1 - q_t) = 1$, hence for $q^m = 1$ it holds that $\tau^m = 0$ i.e., $d^m(1) = 0$.

(iii) $d^m(q^m)$ is strictly decreasing in q^m.

We see from (19), (20) that $\frac{\partial H'^{-1}((1-q_t)K_1)}{\partial q_t} < 0$ and $\frac{\partial H'^{-1}(q_t K_2)}{\partial (1-q_t)} < 0$ for all q_t. Since by the Inada Assumptions $H''(\cdot) > 0, H'(\cdot) \geq 0$ and we have established that $H'^{-1}(\cdot)$ is continuously defined it follows that $d'^m(q^m) < 0$ for all $q^m \in (0,1)$.

The rest of the proof follows directly from Bisin and Verdier (2001) Proposition 1. Inserting $\tau^a = d^i(q_t)$ and $\tau^b = d^i(1 - q_t)$ into (3) and taking the continuous time limit and denoting the continuous rate of change $\dot{q}$ we attain equation (3) in Bisin and Verdier (2001) on page 303:[48]

$$\dot{q} = q(1 - q)[d^a(q) - d^b(1 - q)] \tag{21}$$

We have from part (i) of the proof that $\tau^a > \tau^b$ when $q_t < q^*(\pi)$ and vice versa, hence it follows that $d^a(1 - q) - d^b(q) > 0$ for $q_t < q^*(\pi)$ and $d^a(1 - q) - d^b(q) < 0$ for $q > q^*(\pi)$. Similarly we have from Lemma 3 that $d^a(1 - q^*(\pi)) = d^b(q^*(\pi))$. Note

---

[48]To see why the result is also valid for a discrete time case the see discussion in Bisin and Verdier (2001) in footnote 9 page 303.

that $\frac{\partial \dot{q}_t}{\partial q}\Big|_{q=0} = d^a(0) - d^b(1) > 0$ and $\frac{\partial \dot{q}}{\partial q}\Big|_{q=1} = d^b(0) - d^a(1) > 0$ since $d^m(\cdot)$ satisfies cultural substitution. Since $q^*(\pi)$ is unique, and (21) continuously maps from q into $\dot{q}$, implies that the basin of attraction of $q^*(\pi')$ under π' is $(0, 1)$, which implies that $q_t \to q^*(\pi')$ (Bisin and Verdier, 2001).

$\square$

7.5. Proof of Lemma 5

Lemma 5: Coercion is marginally effective at the beginning and at the end of $[0, \pi_{max}]$, and there is at least one level of coercion, $\hat{\pi}$, that is strictly marginally ineffective; $q^{*\prime}(\hat{\pi}) > 0$.

Proof: Marginal effective is defined as $q^{*\prime}(\pi) < 0$. We show that Assumption 5 implies marginal effectiveness is negative at $\hat{\pi}$, and that marginal effectiveness of coercion is positive at $\pi \in \{0, \pi_{max}\}$. We first show that $q^{*\prime}(\pi) < 0$ for $\pi \in \{0, \pi_{max}\}$. Generally $q^{*\prime}(\pi)$ is given by:

$$q^{*\prime}(\pi) = \frac{(C'(\pi) - 2)\Delta u - C(\pi) + \pi C'(\pi)}{(C(\pi) + 2\Delta u)^2} \tag{22}$$

Assumption 6 implies $C'(\pi) = 0$ at $\pi \in \{0, \pi_{max}\}$. Inserting $C'(\pi) = 0$ into (19) gives:

$$q^{*\prime}(\pi) = \frac{-1}{C(\pi) + 2\Delta u} < 0 \tag{23}$$

We now show that $q^{*\prime}(\hat{\pi}) > 0$. By the functional form assumption on (13) $C''(\hat{\pi}) = 0$ and by the C^2 assumption on $C(\pi)$ in (12) there must be an open interval of ϵ length around $\hat{\pi}$ where $C''(\pi) = 0$. In this interval the function is linear, hence we can assume the coercion resentment function is $C(\pi) = K_0 + K_1\pi$ for some K_1. For

75

$C(\pi) = K_0 + K_1\pi$ marginal effectiveness of coercion in π', $q^{*\prime}(\pi')$, is given by:

$$q^{*\prime}(\pi') = \frac{(K_1 - 2)\Delta u}{(K_1\pi' + 2\Delta u)^2} \tag{24}$$

Since the functional form assumptions on $C(\pi)$ made in (13) implies that $C'''(\pi)$ is strictly non-zero in $[0, \hat{\pi})$ and negative in $(\hat{\pi}, \pi_{max}]$, it must be that $C'(\pi)$ is at its maximum value at $\hat{\pi}$. By the last part of Assumption 6 at least one $\pi' \in (0, \pi_{max})$ must have $C'(\pi) > 2$ and the maximum of $C'(\pi)$ is at $\hat{\pi}$. Hence, it must hold that $C'(\hat{\pi}) > 2$. Inserting $K_1 = 2$ into (20) we attain $q^{*\prime}(\pi) = 0$, and since $K_1 = C'(\hat{\pi}) > 2$ it must hold that $q^{*\prime}(\hat{\pi}) > 0$.

□

7.6. Proof of Lemma 6

Lemma 6: $q^*(\pi)$ is characterized by the following properties:

I) A unique global or local maximum($\pi_{\bar{q}}$) and a unique global minimum($\pi_{\underline{q}}$).

or

II) A unique global or local maximum($\pi_{\bar{q}}$), a local minimum ($\pi_{\underline{q}}$), and a global, potentially unique, minimum ($\pi' \in [\pi_{\underline{q}}^e, \pi_{max}]$).

or

III) A global minimum ($\pi'' \in (0, \hat{\pi})$, where $q^*(\pi'') = 0$).

In addition, there will always be a local or unique global maximum at $q^*(0) = \frac{1}{2}$.

Proof : We show the Lemma by (*i*) demonstrating existence of the extremal points when $q^*(\pi) > 0$ for all $\pi \in [0, \pi_{max}]$. We then show that the classes are exhaustive of all scenarios by first (*ii*) noting what the sign of the derivative of $q^{*\prime}(\pi)$ over $[0, \pi_{max}]$ must be, we then use this to show (*iii*) that any possible $q^*(0) + \int_0^{\pi_{max}} q^{*\prime}(\pi)d\pi$ will place the functional form within either class I), II) or III). Finally, we show (iv) the

76

uniqueness properties of the extremal points.

(i) Existence of extremal points

Suppose $q^*(\pi) > 0$ for all $\pi \in [0, \pi_{max}]$. We show that this implies that there exists an unique minimum in $(0, \hat{\pi})$, $\pi_{\underline{q}}$, and a unique maximum in $(\hat{\pi}, \pi_{max})$, $\pi_{\overline{q}}$, where $\hat{\pi}$ is the turning point of $C''(\pi)$.

We start by showing there is an unique minimum in $[0, \hat{\pi})$. First, we show there exists at least one π such that $q^{*\prime}(\pi) = 0$ in $[0, \hat{\pi})$. We note from the proof of Lemma 5 that:

$$q^{*\prime}(\pi) = \frac{(C'(\pi) - 2)\Delta u + \pi C'(\pi) - C(\pi)}{(C(\pi) + 2\Delta u)^2} \tag{25}$$

It follows from Lemma 5 that $q^{*\prime}(0) < 0$, which by (25) implies $(C'(0) - 2)\Delta u < C(0)$. Similarly it follows from Lemma 5 that $q^{*\prime}(\hat{\pi}) > 0$ which by (25) imply $(C'(\hat{\pi}) - 2)\Delta u + \hat{\pi} C'(\hat{\pi}) > C(\hat{\pi})$. All functions are continuously defined by the C^2 assumption on $C(\pi)$ in (12), hence $(C'(\pi) - 2)\Delta u + \pi C'(\pi)$ and $C(\pi)$ must cross in $(0, \hat{\pi})$, giving $q^{*\prime}(\pi) = 0$ for some $\pi \in [0, \hat{\pi})$. We denote this π value $\pi_{\underline{q}}$. We now show that $\pi_{\underline{q}}$ is a unique value. Note that, following from Assumption 1, (12) and (13), we have that $\Delta u > 0$, $C''(\pi) > 0$ for $\pi \in [0, \hat{\pi})$, and $C'(\pi) > 0$. This implies that the derivative of $(C'(\pi) - 2)\Delta u + \pi C'(\pi)$, $C''(\pi)\Delta u + C'(\pi) + \pi C''(\pi)$ is strictly larger than the derivative of $C(\pi)$, $C'(\pi)$, for all $\pi \in [0, \hat{\pi})$. This implies $C(\pi)$ and $(C'(\pi) - 2)\Delta u + \pi C'(\pi)$ can only cross once in $[0, \hat{\pi})$ and consequently $\pi_{\underline{q}}$ is unique. Finally, we show that the unique $q^{*\prime}(\pi_{\underline{q}}) = 0$ in $[0, \hat{\pi})$ is a minimum. Note that the derivative of $C(\pi)$ is always smaller than the derivative of $(C'(\pi) - 2)\Delta u + \pi C'(\pi)$. Considering (25) we see that this, $q^{*\prime}(0) > 0$, $q^{*\prime}(\hat{\pi}) < 0$ and $0 < \hat{\pi}$ implies that $(C'(\pi) - 2)\Delta u + \pi C'(\pi)$ starts from an initial lower value at $\pi = 0$, surpasses $C(\pi)$

at $\pi_{\underline{q}}$, and is strictly larger than $C(\pi)$ for $\pi \in (\pi_{\underline{q}}, \hat{\pi}]$. Considering (25) we see this implies $q^{*\prime\prime}(\pi) > 0$ for $\pi \in (\pi_{\underline{q}}, \hat{\pi}]$, $q^{*\prime}(\pi) > 0$ for any $\pi > \pi_{\underline{q}}$ and $q^{*\prime}(\pi) < 0$ for any $\pi < \pi_{\underline{q}}$. Hence, $\pi_{\underline{q}}$ is a unique minimum in $[0, \hat{\pi})$.

We now show existence of a unique maximum point in $(\hat{\pi}, \pi_{max}]$. Note that it follows from (12) and (13) that $C''(\pi) < 0$ for $\pi \in (\hat{\pi}, \pi_{max}]$, $C(0) \geq 0$ and $C'(\pi) > 0$. This implies that $q''^*(\pi) = \frac{C''(\pi)(\Delta u + \pi) - C'(\pi)2(C(\pi) + 2\Delta u)}{(C(\pi) + 2\Delta u)^4} < 0$ for $\pi \in (\hat{\pi}, \pi_{max}]$. From Lemma 5 it follows that $q^{*\prime}(\hat{\pi}) > 0$ and $q^{*\prime}(\pi_{max}) < 0$. Hence, $q^{*\prime}(\pi)$ is continuous strictly decreasing in $\pi \in (\hat{\pi}, \pi_{max}]$ from strictly positive to strictly negative, hence there must be one, and only one, $\pi' \in (\hat{\pi}, \pi_{max})$ such that $q^{*\prime}(\pi') = 0$. This π' is defined as $\pi_{\overline{q}}$, the unique maximum in $(\hat{\pi}, \pi_{max}]$.[49]

(ii) The sign of $q^{*\prime}(\pi)$

First, note that from part (i) of the proof we have that $\pi_{\underline{q}} < \hat{\pi} < \pi_{\overline{q}}$. From Lemma 5 and part (i) of the proof it follows that $q^{*\prime}(\pi)$ is strictly increasing from $q^{*\prime}(0) < 0$ to $q^{*\prime}(\pi_{\underline{q}}) = 0$ and onward to $q^{*\prime}(\hat{\pi}) > 2$, and strictly decreasing from $q^{*\prime}(\hat{\pi}) > 2$ to $q^{*\prime}(\pi_{\overline{q}}) = 0$ and onward to $q^*(\pi_{max}) < 0$. Hence, if $q^*(\pi) > 0$ for all $\pi \in [0, \pi_{max}]$, then we note that:

$$q^{*\prime}(\pi) < 0 \text{ for all } \pi \in [0, \pi_{\underline{q}})$$

$$q^{*\prime}(\pi) > 0 \text{ for all } \pi \in (\pi_{\underline{q}}, \pi_{\overline{q}})$$

$$q^{*\prime}(\pi) < 0 \text{ for all } \pi \in (\pi_{\overline{q}}, \pi_{max}]$$

---

[49]Since $\lim_{\pi_{max} \to \infty} \frac{\Delta u - \pi_{max} + C(\pi_{max})}{2\Delta u + C(\pi_{max})} < 1$ for all $\Delta u \in [0, \infty)$ the exterior $q^*(\pi) = 1$ can never be reached; it consequently holds that $q^*(\pi_{\overline{q}}) \in (0, 1)$.

(iii) $q^*(\pi)$ will be characterized by functional form class I), II) or III).

We first show that if $\pi_{\underline{q}}$ is not defined then functional from is of class III). We than show that if $\pi_{\underline{q}}$ is defined it implies that $q^*(\pi)$ is characterized by class I) or class II). We then establish when $q^*(\pi)$ is characterized by class I) or class II).

Note that from part (i) of the proof we have that $q^{*\prime}(0) < 0$ and $q^{*\prime\prime}(0) > 0$. If $q^*(\pi'') = 0$ for some π'' in the interval of $[0, \hat{\pi})$ where $q^{*\prime}(\pi) < 0$ such that $q^*(0) + \int_0^{\pi''} q^{*\prime}(\pi)d\pi = 0$, then since $q^{*\prime}(\pi_{\underline{q}}) = 0$ and $q^{*\prime}(\pi_{\underline{q}}) > 0$ by definition, $\pi_{\underline{q}}$ is not defined. Then $q^*(\pi)$ has a global minimum at this π'' and the functional form is of class III).

If no $\pi'' \in [0, \hat{\pi}]$ while $q^{*\prime}(\pi) < 0$ such that $q^*(\pi'') = 0$ exits, a $q^{*\prime}(\pi') = 0$ where $q^*(\pi') > 0$ exists, this π' is $\pi_{\underline{q}}$. Since $q^{*\prime}(\pi) > 0$ for $(\pi_{\underline{q}}, \pi_{\overline{q}})$ $q^*(\pi) > 0$ for all $\pi \in [0, \hat{\pi}]$ it then follows from part (i) of the proof that this implies there will exists $\pi_{\underline{q}} \in (0, \hat{\pi})$ and $\pi_{\overline{q}} \in (\hat{\pi}, \pi_{max})$.

We note that once $q^*(\pi) = 0$, the SSE for any π is zero. Thus $q^*(\pi)$ ceases to change with π once it reaches 0. Hence, we can impose $q^{*\prime}(\pi) = 0$ for any $q^*(\pi) = 0$ such that we can define integrals of $q^{*\prime}(\pi)$ for $\pi \in [0, \pi_{max}]$ even if $q^*(\pi) = 0$ for some $\pi \in [0, \pi_{max}]$. Hence, we can write any functional form of $q^*(\pi)$ where $\pi_{\underline{q}}$ and $\pi_{\overline{q}}$ is defined as:

$$q^*(0) + \int_0^{\pi_{\underline{q}}} q^{*\prime}(\pi)d\pi + \int_{\pi_{\underline{q}}}^{\pi_{\overline{q}}} q^{*\prime}(\pi)d\pi + \int_{\pi_{\overline{q}}}^{\pi_{max}} q^{*\prime}(\pi)d\pi \qquad (26)$$

Where we know the sign of $q^{*\prime}(\pi)$ in each interval from part (ii) of the proof. Since $\pi_{\underline{q}}$ is defined it follows that $q^*(0) + \int_0^{\pi_{\underline{q}}} q^{*\prime}(\pi)d\pi > 0$.

We note the definition of $\pi_{\underline{q}}^e$, is that $\pi_{\underline{q}} < \pi_{\underline{q}}^e$ and $q^*(\pi_{\underline{q}}^e) \equiv q^*(\pi_{\underline{q}})$. If $\int_{\pi_{\underline{q}}}^{\pi_{\overline{q}}} q^{*\prime}(\pi)d\pi +$ $\int_{\pi_{\overline{q}}}^{\pi_{max}} q^{*\prime}(\pi)d\pi \leq 0$ then, since all functions are continuous, $q^*(\pi_{\underline{q}}^e)$ must be defined. Considering (26) and the part (ii) of the proof we see this implies $q^*(\pi)$ has two minimums, $\pi_{\underline{q}}$ and some $\pi' : \pi' \geq \pi_{\underline{q}}^e$, and one interior maximum, $\pi_{\overline{q}}$. This implies functional form falls within class II).

If $\int_{\pi_{\underline{q}}}^{\pi_{\overline{q}}} q^{*\prime}(\pi)d\pi + \int_{\pi_{\overline{q}}}^{\pi_{max}} q^{*\prime}(\pi)d\pi > 0$, then there will be no $q^*(\pi')$ where $\pi' > \pi_{\underline{q}}$, i.e., $\pi_{\underline{q}}^e$ is not defined. Considering (26) we see this implies all $\pi' > \pi_{\underline{q}}$ has the property $q^*(\pi') > q^*(\pi_{\underline{q}})$ i.e., $q^*(\pi)$ has only one minimum, $\pi_{\underline{q}}$, and one interior maximum, $\pi_{\overline{q}}$. This implies functional falls within class I).

(iv) Properties of the extremal points.
Following from the lemma and the sign of $q^{*\prime}(\pi)$ noted in part (ii) of the proof, there are five possible extremal points, two maximums $\pi \in \{0, \pi_{\overline{q}}\}$, and three possible minimum points, $\pi \in \{\pi', \pi_{\underline{q}}, \pi''\}$ where $\pi' \in [\pi_{\underline{q}}^e, \pi_{max}]$ and $\pi'' \in (0, \pi_{\underline{q}})$. We here establish the properties of the points of importance in the Lemma: $\pi \in \{\pi', \pi_{\underline{q}}, \pi_{\overline{q}}, \pi''\}$.

We first show the properties of $\pi'' \in (0, \pi_{\underline{q}})$. If follows from part (iii) of the proof that if π'' is defined it implies $q^*(\pi'') = 0$ hence π'' is always a global minimum.

We now show when $q^*(\pi_{\overline{q}})$ is a local or global maximum. We have already established in part (i) that $\pi_{\overline{q}}$ is the only interior maximum point. From the sign of $q^{*\prime}(\pi)$ over $[0, \pi_{max}]$ noted in part (ii) of the proof it follows that the other possible maximum point lies at $q^*(0)$. If $q^*(0) < q^*(\pi_{\overline{q}})$, $\pi_{\overline{q}}$ is a unique global maximum, if $q^*(0) \geq q^*(\pi_{\overline{q}})$ then $q^*(\pi_{\overline{q}})$ is a local maximum.

We now show when $q^*(\pi_{\underline{q}})$ is a unique global, non-unique global or local minimum. Suppose that $\int_{\pi_{\underline{q}}}^{\pi_{\overline{q}}} q^{*\prime}(\pi)d\pi + \int_{\pi_{\overline{q}}}^{\pi_{max}} q^{*\prime}(\pi)d\pi > 0$. From (iii) of the proof this implies functional form is of class I), and $\pi_{\underline{q}}$ is the only minimum and hence a unique global minimum point. Suppose $\int_{\pi_{\underline{q}}}^{\pi_{\overline{q}}} q^{*\prime}(\pi)d\pi + \int_{\pi_{\overline{q}}}^{\pi_{max}} q^{*\prime}(\pi)d\pi = 0$, then $\pi_{\underline{q}}$ is a non-unique global minimum, since it then must hold that $q^*(\pi_{\underline{q}}) = q^*(\pi_{max})$. Suppose, $\int_{\pi_{\underline{q}}}^{\pi_{\overline{q}}} q^{*\prime}(\pi)d\pi + \int_{\pi_{\overline{q}}}^{\pi_{max}} q^{*\prime}(\pi)d\pi < 0$ then $\pi_{\underline{q}}$ is a non-unique local minimum, since this implies there exists an π' such that $q^*(\pi_{\underline{q}}) < q^*(\pi')$.

We now show $q^*(\pi')$ where $\pi' \in [\pi_{\overline{q}}^e, \pi_{max}]$, is a global minimum. Since $q^{*\prime}(\pi) < 0$ for $\pi \in (\pi_{\underline{q}}, \pi_{max})$ as established in (ii) of the proof, this minimum is unique global if $\pi' = \pi_{max}$ and $q^*(\pi') < q^*(\pi_{\underline{q}})$. The minimum π' is non-unique global if $\pi' < \pi_{max}$, this implies $q^*(\pi'''') = 0$ for all $\pi'''' \le \pi'$. The minimum π' is also non-unique global if $q^*(\pi') = q^*(\pi_{\underline{q}})$ and $\pi' = \pi_{max}$ as follows from the preceding discussion of the properties of $\pi_{\underline{q}}$.

$\square$

7.7. Proof of Proposition 1.

Proposition 1: Let π' denote a level of coercion such that $q^*(\pi') = 0$. The optimal level of coercion under no constraint, π^{NC}, will be as follows for the different classes of $q^*(\pi)$:

I) $\pi^{NC} = \pi_{\underline{q}}$.

II) $\pi^{NC} = \pi' \in (\pi_{\underline{q}}^e, \pi_{max})$, if no π' is defined then $\pi^{NC} = \pi_{max}$.

III) $\pi^{NC} = \pi' < \pi_{\underline{q}}$, where π' is always defined.

Proof: This proof follows from Lemma 6 and the assumption that the authority is minimizing $q^*(\pi)$, as captured in (7).

□

7.8. Proof of Corollary of Proposition 1.

Corollary of Proposition 1: An authority facing no constraint on coercion will only restrain its use of coercion when SSE $q^*(\pi)$, is of class I).

Proof: A restraint on coercion implies a coercion level π' being strictly smaller than the highest implementable coercion level with a non-zero $q^*(\pi)$, it follows directly from Proposition 1 that this only occurs under $q^*(\pi)$ of class I).
□

7.9. Proof of Proposition 2.

Proposition 2: If a constraint affects the coercion use under an exogenous constraint, $\rho \leq \pi^{NC}$ and $\rho \neq \pi_q^e$, it holds that:
(i) $\pi^{EC} = \rho$ if and only if $\rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.
(ii) $\pi^{EC} = \pi_{\underline{q}} < \rho$ if and only if $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.

Proof: We first note the three possible scenarios of π^{EC} of $\rho \in [0, \pi_{max}]$ and then demonstrate that claim (i) and (ii) of the proposition.

We note that it follows from that the authority is minimizing $q^*(\pi)$ by (7) and Lemma 6 that if $\rho \neq \pi_q^e$ and $\rho \leq \pi^{NC}$ there are three different scenarios of $\rho \in [0, \pi_{max}]$:

I) $\rho \in [0, \pi_{\underline{q}}] \rightarrow \pi^{EC} = \rho$.
From part (ii) of proof of Lemma 6, it holds that $q^{*\prime}(\pi) < 0$ for all $\pi \in [0, \pi_{\underline{q}})$. Hence the minimal $q^*(\pi)$ for $\rho \in (0, \pi_{\underline{q}})$ is always equal to ρ.

II) $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e) \to \pi^{EC} = \pi_{\underline{q}} < \rho$.

By the proof of Lemma 6 part (iii) $\pi_{\underline{q}}$ is the minimum value in $[0, \pi_{max}]$ unless $\pi_{\underline{q}}^e$ is defined. By definition $\pi_{\underline{q}}^e$ is a unique π value larger than $\pi_{\underline{q}}$ such that $q^*(\pi_{\underline{q}}^e) = q^*(\pi_{\underline{q}})$ which follows from (26) and the proof of Lemma 6 part (ii). Hence, for every $\pi' \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ it holds that $q^*(\pi') > q^*(\pi_{\underline{q}})$ and $\pi_{\underline{q}}$ must be the minimum of the open interval of $[0, \pi_{\underline{q}}^e)$.

III) $\rho \in (\pi_{\underline{q}}^e, \pi_{max}] \to \pi^{EC} = \rho$.

From part (ii) of proof of Lemma 6 $q^{*\prime}(\pi) < 0$ for all $\pi \in (\pi_{\overline{q}}, \pi_{max}]$; since $\pi_{\underline{q}}^e \in (\pi_{\overline{q}}, \pi_{max})$ any $\pi' > \pi_{\underline{q}}^e$ implies $q^*(\pi_{\underline{q}}^e) > q^*(\pi')$. By Lemma 6 it follows that $\pi = \pi_{\underline{q}}$ is the minimum of $\pi \in [0, \pi_{\underline{q}}^e)$. Since $q^*(\pi_{\underline{q}}) \equiv q^*(\pi_{\underline{q}}^e)$ by definition in (14), the minimum $q^*(\pi)$ when choosing a $\pi^{EC} \in [0, \rho]$ where $\rho \in (\pi_{\underline{q}}^e, \pi_{max}]$ is ρ.

Note that $\rho \in [0, \pi_{\underline{q}}]$ or $\rho \in (\pi_{\underline{q}}^e, \pi_{max}]$ implies $\rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$. Thus I) and III) can be combined so that the different scenarios of $\rho \in [0, \pi_{max}]$ can be stated:

$$\rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e) \to \pi^{EC} = \rho \tag{27}$$

$$\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e) \to \pi^{EC} = \pi_{\underline{q}} < \rho \tag{28}$$

Note that the proposition states that $\rho \neq \pi_{\underline{q}}^e$ and I) implies that $\rho = \pi_{\underline{q}} \to \pi^{EC} = \pi_{\underline{q}} = \rho$. This, (28) and (27) covers all possible scenarios of π^{EC} for $\rho \in [0, \pi_{max}]$ which implies that:

$$\pi^{EC} = \rho \to \rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e) \tag{29}$$

$$\pi^{EC} = \pi_{\underline{q}} < \rho \to \rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e) \tag{30}$$

83

Part (i) of the proposition follows from (27) and (29). Part (ii) of the proposition follows from (28) and (30).

□

7.10. Proof of Corollary of Proposition 2.

Corollary of Proposition 2: An authority will restrain its use of coercion as a response to a constraint if and only if the $q^*(\pi)$ is of class II) and the constraint is in the inefficient interval of coercion; $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$.

Proof: We first note that by definition on page 28 an authority exhibiting restraint as a response to the constraint imposes a π^{EC} that is:

1. A response to a constraint; a π^{EC} different than its optimal adjustment without constraints, $\pi^{EC} \neq \pi^{NC}$.

2. A restraint; a π^{EC} strictly lower than its highest implementable level, $\pi^{EC} < \rho \leq \pi_{max}$.

We first show the if part, that $q^*(\pi)$ is of class II) and $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ implies a restrain on coercion as a response to a constraint. We then show the only if part by first demonstrating that if $\rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ then there is no restrain on coercion. Finally, we show that if $q^*(\pi)$ of class I) or class III) and $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ then π^{EC} is not a response to a constraint.

If $q^*(\pi)$ is in class II) and $\rho \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$, then from part ($ii$) of Proposition 2 $\pi^{EC} = \pi_{\underline{q}} < \rho$. Then π^{EC} is then a restraint as a response to a constraint since $\pi^{EC} = \pi_{\underline{q}} < \pi_{\underline{q}}^e < \pi^{NC}$.

If $\rho \notin (\pi_{\underline{q}}, \pi_{\bar{q}}^e)$ and $\rho \neq \pi_{\underline{q}}^e$, then $\pi^{EC} = \rho$ by part $i)$ of Proposition 2, hence π^{EC} is not a restrain. If $\pi^{NC} = \pi^{EC} = \{\pi_{\bar{q}}^e, \pi_{\underline{q}}\}$, then π^{EC} is not a response to a constraint.

If $q^*(\pi)$ is of class I) and $\rho \in (\pi_{\underline{q}}, \pi_{\bar{q}}^e)$, then from Proposition 1 and 2 we have that $\pi^{NC} = \pi^{EC} = \pi_{\underline{q}}$, thus π^{EC} is not a response to a constraint. If $q^*(\pi)$ is of class III) and $\rho \in (\pi_{\underline{q}}, \pi_{\bar{q}}^e)$, then by Proposition 1 and 2 $\pi^{NC} = \pi^{EC} = \pi' < \rho$ where $q^*(\pi') = 0$, consequently π^{EC} is not a response to a constraint.

$\square$

7.11. Proof of Proposition 3.

Proposition 3: Weak coercion reliance occurring without strong coercion reliance implies a strictly marginally inefficient interval of coercion between intervals of marginal effective coercion levels.

Proof: Strong coercion reliance implies $\pi = 0$ being an unsustainable coercion level. Weak coercion reliance implies there exists $\pi' < \pi_0$ such that $\pi' \in [0, \pi_0)$ is unsustainable. Weak coercion reliance without strong coercion reliance implies $\pi = 0$ being an sustainable level of coercion, $0 \geq \rho(q^*(0))$, while there exists some level $0 < \pi' < \pi_0$ that is unsustainable i.e., $\rho(q^*(\pi')) < \pi'$. Since π_0 by definition implies an interior SSE not breaching the insurrection constraint and by Assumption 7 $\rho'(\cdot) < 0$, weak coercion reliance occurring without strong coercion reliance must imply that there exists a π' such that $\rho(q^*(\pi')) < \rho(q^*(0)) \leq \rho(q^*(\pi_0))$. Since $\rho'(\cdot) < 0$, this implies that $q^*(\pi') > q^*(\pi_0)$ where $\pi' < \pi_0$.

$\pi' < \pi_0$ while $q^*(\pi') > q^*(\pi_0) \geq q^*(0)$ cannot occur without an interval of π such that $q^{*\prime}(\pi) < 0$ in between intervals of π such that $q^{*\prime}(\pi) > 0$, which is by the definition

of marginal efficient on page 21, a marginal inefficient interval preceded and followed by marginal efficient levels of coercion.

$\square$

7.12. Proof of Proposition 4.

Proposition 4: For any initial condition, π_0, the dynamically stable equilibrium π^{IC} is a coercion level equal to either:

I) The unconfrontational level of coercion as an interior point of $\mathbf{S}_{\pi_0}$: $\pi^{IC} = \pi_{\underline{q}}$.

II) A strategic constraint at the upper bound of $\mathbf{S}_{\pi_0}$: $\pi^{IC} = \overline{\pi_{fix}}$.

III) The upper feasibility constraint at the bound of $\mathbf{S}_{\Pi}$: $\pi^{IC} = \pi_{max}$.

Proof: We show that the π' corresponding to any minimum point of any $\mathbf{Q}_{\pi_0}$, which by definition is equal to π^{IC}, will fall under either case I), II) or III), hence the proposition.[50] First note that trivially any minimum point in $\mathbf{Q}_{\pi_0}$ must correspond to a π^{IC} in the interior of a subset of an $\mathbf{S}_{\pi_0}$, $\mathbf{s}$, or at boundary of an $\mathbf{s}$.

Suppose the minimum of $\mathbf{Q}_{\pi_0}$ corresponds to an interior point in an $\mathbf{s}$. As established in Lemma 6, $q^*(\pi)$ has at most one interior minimum point, $\pi_{\underline{q}}$, and since $\rho'(q^*(\pi)) < 0$ always holds, $q^{*\prime}(\pi) = 0$ must hold at a minimum of $\mathbf{Q}_{\pi_0}$ corresponding to an interior minimum of $\mathbf{s}$. Thus π^{IC} must be equal to $\pi_{\underline{q}}$ and π^{IC} fall under case I).

Suppose the minimum of $\mathbf{Q}_{\pi_0}$ corresponds to a π^{IC} that is the limit of a subset $\mathbf{s}$ and this limit is different from π_{max}. π^{IC} must be at an upper limit of $\mathbf{s}$, since at

---

[50]If there are several infimum points, any will correspond to a dynamically stable equilibrium, as the authority will not have any incentive to change π.

lower thresholds of $\mathbf{s}$ lowering π increases $q^*(\pi)$, as follows from the proof of Proposition 3. Since the limit π^{IC} is an upper limit different from π_{max} it implies there exists a $\pi^{IC} < \pi' < \pi_{max}$, such that π' is infinitesimally larger than the upper limit of the subset. Since $\pi' \notin \mathbf{s}$ it implies that $\rho(q^*(\pi')) > \pi'$. Since $\rho(\cdot)$ is assumed to be a continuous mapping with a continuous derivative, it cannot discontinuously jump from π^{IC}, over or on the 90-degree fix-point-line, to a point π' under the line, without crossing the fix-point-line.[51] Hence the minimum of $\mathbf{Q}_{\pi_0}$ must correspond to an upper limit on the fix-point-line $\pi^{IC} = \overline{\pi_{fix}}$ which falls under case II).

Suppose the minimum of $\mathbf{Q}_{\pi_0}$ corresponds to an upper limit of a subset $\mathbf{s}$ and this limit is $\pi^{IC} = \pi_{max}$ and corresponding to case III).

□

7.13. Proof of Proposition 5.

Proposition 5: If and only if there exist initial conditions, $\overline{\pi}_0 \neq \underline{\pi}_0$, such that the set of implementable coercion levels from $\overline{\pi}_0$ or $\underline{\pi}_0$ differ, $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$, will different initial conditions $\pi_0 = \overline{\pi}_0$ and $\pi_0 = \underline{\pi}_0$ give different dynamically stable equilibria; $\overline{\pi}^{IC}_{\overline{\pi}_0} \neq \underline{\pi}^{IC}_{\underline{\pi}_0}$.

Proof: We first show that if $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$ then $\overline{\pi}^{IC}_{\overline{\pi}_0} \neq \underline{\pi}^{IC}_{\underline{\pi}_0}$. We then show if $\overline{\pi}^{IC}_{\overline{\pi}_0} \neq \underline{\pi}^{IC}_{\underline{\pi}_0}$ then $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$.

Suppose $\overline{\pi}_0 > \underline{\pi}_0$ and $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$. By definition of the set of implementable coercion levels there must be at least one π' such that $\pi' \notin \underline{\mathbf{S}}_{\underline{\pi}_0}$ but $\pi' \in \overline{\mathbf{S}}_{\overline{\pi}_0}$, since

---

[51] The fix-point-line for $\hat{\rho}(\pi)$ is illustrated in Figure. 2 on page 34.

if this was not the case then the sets would be the same sets; i.e., this is implied by $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$. This implies that $q^*(\pi') < inf\{\mathbf{Q}_{\underline{\pi}_0}\}$, as $\rho(q^*(\pi))$ is monotonically increasing in $q^*(\pi)$ and π' cannot be reached from $\underline{\pi}_0$. Suppose that the difference between the sets consist of this single coercion level π'. This $\pi' > sup\{\mathbf{S}_{\underline{\pi}_0}\}$ must then be equal to $\overline{\pi}^{IC}$, since π' must corresponds to $inf\{\mathbf{Q}_{\overline{\pi}_0}\}$. This π' is different from $\underline{\pi}^{IC}$ since π' is not in $\mathbf{S}_{\underline{\pi}_0}$.[52]

Suppose the dynamically stable equilibria are different and that $\overline{\pi}^{IC}_{\overline{\pi}_0} > \underline{\pi}^{IC}_{\underline{\pi}_0}$.[53] By definition $\overline{\pi}^{IC}_{\overline{\pi}_0}$ is corresponding to $inf\{\mathbf{Q}_{\overline{\pi}_0}\}$. Since $\rho'(\cdot) < 0$ this implies that it must hold that $q^*(\overline{\pi}^{IC}_{\overline{\pi}_0}) < q^*(\underline{\pi}^{IC}_{\underline{\pi}_0})$ since $\underline{\pi}^{IC}_{\underline{\pi}_0}$ can be implemented from $\overline{\pi}^{IC}_{\overline{\pi}_0}$, but $\overline{\pi}^{IC}_{\overline{\pi}_0}$ gives a lower $q^*(\pi)$ than $\underline{\pi}^{IC}_{\underline{\pi}_0}$ by definition. Hence, there must be at least one π' such that $\pi' \notin \underline{\mathbf{S}}_{\underline{\pi}_0}$ but $\pi' \in \overline{\mathbf{S}}_{\overline{\pi}_0}$, namely $\overline{\pi}^{IC}_{\overline{\pi}_0}$. By definition of the set of implementable coercion levels this implies $\overline{\mathbf{S}}_{\overline{\pi}_0} \triangle \underline{\mathbf{S}}_{\underline{\pi}_0} \neq \emptyset$.

□

8. Appendix 2: Linear coercion resentment functions

Assuming the coercion resentment function to be linear; $C(\pi) = K_0 + K_1\pi$ for some K_0, K_1 yields:

$$\frac{\partial q^*(\pi)}{\partial \pi} = \frac{(K_1 - 2)\Delta u}{(K_1\pi + 2\Delta u)^2} \tag{31}$$

---

[52]Note that in the special case where there are two dynamically stable equilibria but equal sets of implementable coercion levels then it might be more *likely* to end up in different equilibria due to different initial conditions.

[53]Note that the proposition does not cover the special case of multiple dynamically stable equilibria, $\overline{\pi}^{IC}_{\overline{\pi}_0} = \underline{\pi}^{IC}_{\underline{\pi}_0} = \{\pi_{\overline{q}}, \pi^e_{\overline{q}}\}$.

Using coercion as a means to change $q^*(\pi)$ would simply not be a useful tool for $K_1 = 2$, counterproductive for $K_1 > 2$, or productive at any level for $K_1 < 2$. The solution to the authority's minimization problem in (7) is trivial for linear coercion functions; either always coerce as hard as possible or never coerce at all depending on the coercion resentment, K_1, is greater or smaller than 2. Similarly, for a convex or a concave coercion resentment function the problem of setting the optimal coercion level will have a unique extremal point at the π' that solves:

$$\Delta u = \frac{2C(\pi') - \pi'}{2 + 3C'(\pi')} \tag{32}$$

9. Appendix 3: Further interpretations of the model

9.1. Appendix 3.1: Policy implications of Proposition 2

Proposition 2 has relevant implications for the policy problem of an external agency setting a constraint ρ to limit an authority's' use of coercion when $q^*(\pi)$ is of Class II).

Assume that the cost of enforcing the constraint is $K_0(\pi_{max} - \rho)$, where $K_0 > 0$. Setting a $\tilde{\rho} \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$ will have several benefits relative to a constraint $\rho \notin (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$. Imposing $\tilde{\rho}$ implies a costless reduction of the equilibrium coercion level; since the cost of imposing $\tilde{\rho}$ is $K_0(\pi_{max} - \tilde{\rho})$, and Proposition 2 implies that at $\tilde{\rho}$ the imposed coercion level π^{EC} is $\pi_{\underline{q}}$, the actual reduction of coercion is $(\pi_{max} - \pi_{\underline{q}})$. Hence, the cost of reduction is given by $(\pi_{max} - \tilde{\rho})$ while the actual reduction is $(\pi_{max} - \pi_{\underline{q}})$, implying that the reduction $(\pi_{max} - \pi_{\underline{q}}) - (\pi_{max} - \tilde{\rho}) = (\tilde{\rho} - \pi_{\underline{q}})$ is achieved at no cost. Further, assume that the external agency has imperfect information of where the coercion level lies, giving the authority a possibility to increase π without the external agency being able to identify the increase. At $\pi_{\underline{q}}$ the authority has no incentive to marginally increase π in equilibrium, as this would imply imposing a coercion level

89

π in the inefficient interval i.e., $\pi \in (\pi_{\underline{q}}, \pi_{\underline{q}}^e)$. A final benefit is that the $\tilde{\rho}$ constraint is not binding in equilibrium, which will often reduce its salience.

9.2. Appendix 3.2: Interpretations of the insurrection constraint

One natural way to reason about why the insurrection constraint has $\rho'(q^*(\pi)) < 0$, is to assume that increasing the size of the non-state morality always increases their capability for committing an successful insurrection. The lower threshold for committing an insurrection then follows from a higher probability of a successful outcome of an insurrection. Capability to attain a successful outcome of an insurrection will grow with $q^*(\pi)$ for a wide number of applications; hence the assumption of $\rho'(q^*(\pi)) < 0$.

In applications of the model where military capability determines the capability to preform a successful insurrection, the functional form of $\rho(q^*(\pi))$ is determined by the current military technology's ability to transform the share of a morality individuals, $q^*(\pi)$, into military capability. The derivative of the insurrection constraint function at a particular SSE level, $\rho'(q^*(\pi'))$, will be determined by the relative labor intensity of military power. Assuming the insurrection constraint to be independent of the SSE, $\rho'(q^*(\pi)) = 0$ for all $q^*(\pi) \in (0, 1)$, implies a military technology solely dependent on capital. A constant derivative, $\rho'(q^*(\pi)) = K$, for all $q^*(\pi) \in (0, 1)$, implies the military technology where every individual of the population has equal ability to exert use of military force and there is no scarcity of capital.

Applying the model to a democratic system, the endogenous constraint will reflect a situation where an authority faces an undesirable outcome contingent on the level of $q^*(\pi)$ not moving beyond some threshold needed to issue a forced referendum or a motion of no confidence.

10. Appendix 4: The set of implementable coercion levels

The following discussion provides some conjunctures about how the set of implementable coercion levels would be under insurrections constraints with other mappings between q_t and the threshold level of insurrection, (4.1), and iterative processes for $\mathbf{S}_{\pi_0}$ where π can be set at any t, (4.2).

10.1. Appendix 4.1 : Sufficiency of constraints on $\rho(q^*(\pi))$

We here discuss what requirement must be put on the model to insure the insurrection constraint is not breached given other relations between q_t and the threshold level of insurrection. We then discuss how this changes the set of implementable coercion levels.

The model considers an insurrection constraint on $q^*(\pi)$ rather than q_t. For a solution considering an insurrection constraint on $q^*(\pi)$ to be sufficient to imply that the solution would also hold for an insurrection constraint dependent on q_t, further restrictions are needed. The restrictions must insure that whenever a coercion level, π', satisfying an initial insurrection constraint $\rho(q^*(\pi_0)) \geq \pi'$ is imposed, then this must imply that $\rho(q_t) \geq \pi'$ holds for all q_t in the sequence of q_t values in the convergence sequence from $q^*(\pi_0)$ towards $q^*(\pi')$. Following the notation in Lemma 4 we denote this sequence of q_t values $\{q_t\}_{\pi'}$. We here discuss when the following criterion is met:

$$\text{If } \rho(q^*(\pi_0)) \geq \pi' \text{ and } \rho(q^*(\pi')) \geq \pi' \text{ then } \rho(q_t) \geq \pi' \text{ for all } q_t \in \{q_t\}_{\pi'} \qquad (33)$$

Since $\rho(q^*(\pi))$ is monotonically strictly increasing, $\rho(q_t) \geq \pi'$ for all q_t is insured if no $q_t \in \{q_t\}_{\pi'}$, is larger than the start of the convergence process; i.e., it must hold that $q^*(\pi_0) \geq q_t$, for all $q_t \in \{q_t\}_{\pi'}$. This is equivalent to a requirement of

no-overshooting of $q^*(\pi_0)$ i.e., $q^*(\pi_0) \geq q_t$ for all $q_t \in \{q_t\}_{\pi'}$. Assuming a change in π happening at time $t = 0$, then from (3) and requiring $q_t < q_{t+1}$ for any convergence path from $q^*(\pi_0)$ to $q^*(\pi')$ we derive that the cost function in socialization efforts must be sufficiently bounded for changes within $q_t \in (0, 1)$ such that:

$$\Delta_t \tau^m \geq \Delta_t \tau^m \Delta_{t+1} \tau^m + \Delta_{t+1} \tau^m \rightarrow 1 \geq \Delta_{t+1} \tau^m [1 + \frac{1}{\Delta_t \tau^m}] \qquad (34)$$

If (34) holds for all possible combinations of moving from one $q_t \in (0, 1)$ to another $q_{t+1} \in (0, 1)$ then (33) is satisfied. Hence, the requirement of no-overshooting is fulfilled as long as $|H(\tau_t) - H(\tau_{t+1})|$ is sufficiently bounded for changes in $q \in (0, 1)$.

Assume the insurrection constraint would be dependent on a moving average $\tilde{\rho}(\bar{q}_{N,t})$ where $\bar{q}_{N,t} \equiv \frac{\Sigma_{i=0}^N q_{t-i}}{N+1}$. Further, assume that the convergence process from $q^*(\pi_0)$ to $q^*(\pi')$ in (33) happens within T periods. We know that $q^*(\pi')$ and $q^*(\pi_0)$ is sustainable and it follows from the proof of Lemma 4 part (iii) that any average will converge towards $q^*(\pi')$; i.e., it holds that $\bar{q}_{N,T} \rightarrow q^*(\pi')$ as $N \rightarrow \infty$. Hence, in the model as specified, (33) holds for an infinite moving average i.e., $N = \infty$, infinite inertia. More simply put; (33) holds if the military capability remains as at $q^*(\pi_0)$ throughout the convergence process.

The smaller N, the stricter the requirement is needed on the convergence processes; and for $N = 0$ i.e., no inertia (34) must always hold. As discussed on page 20, we have established that the convergence process will happen with every other generation in the process being above or bellow the SSE $q^*(\pi')$ i.e., $q_t < q_{t+2} < q^*(\pi') < q_{t+3} < q_{t+1}$. Hence, for any inertia process that can be described by a lag of more than two periods, $N \leq 2$, a requirement of the cost function in socialization efforts will be strictly weaker than (34). Further, since a shorter convergence time implies that $\bar{q}_{N,T}$ is

closer to $\bar{q}_{N,0}$, we see that the requirement will be weaker if the convergence process is shorter i.e., if T is low; trivially for convergence in one period, $T = 1$, (33) always holds.

The discussion above has been dealing with considerations of whether a constraint on the SSE can be considered a not sufficiently strict criterion to analyze which $q^*(\pi)$ an authority can dynamically reach. In other words, for another insurrection constraint dependent on q_t there might be $q^*(\pi) \in \mathbf{Q}_{\pi_0}$ the authority might not reach. Further, we have argued that it appears that the set of implementable coercion levels for an insurrection constraint dependent on $\bar{q}_{N,t}$, $\tilde{\mathbf{S}}_{\pi_0,N,T}$ will converge towards $\mathbf{S}_{\pi_0}$, as the inertia of military capability converges to infinity, $N \to \infty$, and the number of generations it takes to convergence between steady states converges to one, $\frac{1}{T} \to 1$.

10.2. Appendix 4.2: The set of implementable coercion levels when the authority can reset π at every t

In the specified model the set of implementable coercion level is given by what the the authority can reach by setting π' in a $q^*(\pi_0)$ and then reset π once $q^*(\pi')$ is reached. Assume, as in Appendix 4.1, that a insurrection constraint is dependent on q_t rather than $q^*(\pi)$, and that π can be reset π at any t in the convergence sequence $\{q_t\}_{\pi'}$, defined in Lemma 4. The authority would then, potentially, be able to reach $q^*(\pi) \notin \mathbf{Q}_{\pi_0}$. This can arise as there might be q_t values in the convergence sequence, $\{q_t\}_{\pi'}$, from which the authority might be able to implement some π'' not implementable in $\mathbf{S}_{\pi_0}$ and thus reach $q^*(\pi) \notin \mathbf{Q}_{\pi_0}$. Investigating what states would then be reachable would require a further inquiry into the extremal values of the

convergence sequence $\{q_t\}_{\pi'}$. Which states that would be sustainable, $\mathbf{S}_\Pi$, would not change and there could still be limits to what is reachable from some initial condition; an authority could still be strategically constrained at an upper bound attractor fix-point $\overline{\pi_{fix}}$. In other words, $\mathbf{S}_{\pi_0}$ might be different for other iterative processes, but it appears that all established results would qualitatively hold.

11. Bibliography

Acemoglu, Daron and Alexander Wolitzky. 2014. "Cycles of conflict: An economic model." *The American Economic Review* 104(4):1350–1367.

Acemoglu, Daron and James A. Robinson. 2017. The Emergence of Weak, Despotic and Inclusive States. Working Paper 23657 National Bureau of Economic Research.

Alesina, Alberto and Bryony Reich. 2013. Nation Building. Technical report National Bureau of Economic Research.

Alesina, Alberto and Enrico Spolaore. 2003. *The Size of Nations.* MIT Press.

Allardt, Erik. 1979. *Implications of the ethnic revival in modern, industrialized society: a comparative study of the linguistic minorities in Western Europe.* Societas Scientiarum Fennica.

Barfield, Thomas J. 2010. *Afghanistan: A Cultural and Political History.* Princeton University Press.

Bennigsen, Alexandre and Chantal Lemercier-Quelquejay. 1967. *Islam in the Soviet Union.* Pall Mall Press.

Bisin, Alberto, Eleonora Patacchini, Thierry Verdier and Yves Zenou. 2011. "Formation and persistence of oppositional identities." *European Economic Review* 55(8):1046–1071.

Bisin, Alberto and Thierry Verdier. 2000. "A model of cultural transmission, voting and political ideology." *European Journal of Political Economy* 16(1):5–29.

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

Bisin, Alberto and Thierry Verdier. 2010. The economics of cultural transmission and socialization. Technical report National Bureau of Economic Research.

Bowles, Samuel and Herbert Gintis. 2011. *A Cooperative Species: Human Reciprocity and its Evolution.* Princeton University Press.

Cantoni, Davide. 2015. "The economic effects of the Protestant Reformation: testing the Weber hypothesis in the German lands." *Journal of the European Economic Association* 13(4):561–598.

Carvalho, Jean-Paul. 2013. "Veiling." *The Quarterly Journal of Economics* 128(1):337–370.

Carvalho, Jean-Paul and Mark Koyama. 2013. Resisting Education. Technical report University Library of Munich, Germany.

Conquest, Robert. 1970. *The Nation Killers: The Soviet Deportation of Nationalities.* Macmillan London.

Fouka, Vasiliki. 2015. "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I." *Working paper* .

Froese, Paul. 2008. *The Plot to Kill God: Findings From the Soviet Experiment in Secularization.* University of California Press.

Greif, Avner. 2008. "The Normative Foundations of Institutions and Institutional Change." *Unpublished manuscript* .

Greif, Avner and Jared Rubin. 2014. "Endogenous Political Legitimacy: The English Reformation and the Institutional Foundations of Limited Government." *Working paper, Stanford University* .

Greif, Avner and Steven Tadelis. 2010. "A theory of moral persistence: Crypto-morality and political legitimacy." *Journal of Comparative Economics* 38(3):229–244.

Helfferich, Tryntje. 2009. *Thirty Years' War: An Anthology of Sources.* Hackett.

Huddy, Leonie, David O Sears and Jack S Levy. 2013. *The Oxford handbook of political psychology.* Oxford University Press.

Johnson, Noel D and Mark Koyama. 2013. "Legal centralization and the birth of the secular state." *Journal of Comparative Economics* 41(4):959–978.

Kula, Marcin. 2005. "Communism as Religion." *Totalitarian Movements and Political Religions* 6(3):371–381.

Laitin, David D. 1998. *Identity in formation: The Russian-speaking populations in the near abroad.* Vol. 22 Cambridge University Press.

Levi, Margaret. 1997. *Consent, dissent, and patriotism.* Cambridge University Press.

Levi, Margaret. 1999. "Death and taxes: extractive equality and the development of democratic institutions." *Democracy's Value* pp. 112–31.

Marshall, Richard H, Thomas E Bird and Andrew Blane. 1971. *Aspects of religion in the Soviet Union, 1917-1967.* University of Chicago Press.

Mendelson, Bert. 1975. *Introduction to topology.* Courier Corporation.

Nash, John. 1953. "Two-person cooperative games." *Econometrica* pp. 128–140.

Nexon, Daniel. 2009. "The struggle for power in early modern Europe." *Princeton University Press* .

North, Douglass C and Robert Paul Thomas. 1973. *The rise of the western world: A new economic history.* Cambridge University Press.

Northrop, Douglas. 2001. "Subaltern dialogues: Subversion and resistance in Soviet Uzbek family law." *Slavic review* pp. 115–139.

Nye, Joseph S. 1967. "Corruption and political development: A cost-benefit analysis." *The American Political Science Review* pp. 417–427.

Onorato, Massimiliano Gaetano, Kenneth Scheve and David Stasavage. 2014. "Technology and the Era of the Mass Army." *The Journal of Economic History* 74(02):449–481.

Petrosyan, Leon A. 2016. *Dynamic Games with Perfect Information.* Springer.

Philpott, Daniel. 2001. *Revolutions in sovereignty: how ideas shaped modern international relations.* Princeton University Press.

Rae, Heather. 2002. *State identities and the homogenisation of peoples.* Vol. 84 Cambridge University Press.

Ringmar, Erik. 2007. *Identity, interest and action: a cultural explanation of Sweden's intervention in the Thirty Years War.* Cambridge University Press.

Roberts, Michael. 1954. *The military revolution, 1560-1660.* Queen's University of Belfast.

Rokkan, Stein. 1999. *State formation, nation-building, and mass politics in Europe: the theory of Stein Rokkan: based on his collected works.* Clarendon Press.

Rubin, Jared. 2014. "Printing and Protestants: an empirical test of the role of printing in the Reformation." *Review of Economics and Statistics* 96(2):270–286.

Sambanis, Nicholas and Moses Shayo. 2013. "Social identification and ethnic conflict." *American Political Science Review* 107(02):294–325.

Schøyen, Øivind. 2011. *Legitimacy maximizing authorities.* Master thesis. The London School of Economics and Political Science.

Sutherland, Nicola M. 1988. The Crown, the Huguenots, and the Edict of Nantes. In *The Huguenot Connection: The Edict of Nantes, Its Revocation, and Early French Migration to South Carolina.* Springer pp. 28–48.

Theibault, John. 1997. "The Demography of the Thirty Years War Re-revisited: Günther Franz and his Critics." *German History* 15(1):1–21.

Tilly, Charles. 1992. *Coercion, capital, and European states, AD 990-1992.* Blackwell Oxford.

Tilly, Charles and Gabriel Ardant. 1975. *The formation of national states in Western Europe.* Vol. 8 Princeton University Press.

Weingast, Barry R and Donald Wittman. 2008. *The Oxford handbook of political economy.* Oxford University Press.

Wilson, Peter Hamish. 2009. *The Thirty Years War: Europe's Tragedy.* Harvard University Press.

Issued in the series Discussion Papers 2016

2016

01/16 January, **Ingvild Almås and** Anders Kjelsrud**,** "Pro-poor price trends and inequality|the case of India".

02/16 January, Tuomas Pekkarinen, **Kjell G. Salvanes,** and Matti Sarvimäki, «The evolution of social mobility: Norway over the 20th century"

03/16 February, Sandra E. Black, Paul J. Devereux, and **Kjell G. Salvanes**, "Healthy(?), Wealthy, and Wise. Birth Order and Adult Health".

04/16 March, Hiroshi Aiura, "The effect of cross-border healthcare on quality, public health insurance, and income redistribution"

05/16 March**, Jan Tore Klovland**, "Shipping in dire straits: New evidence on trends and cycles in coal freights from Britain, 1919-1939"

06/16 April, Branko Bošković and **Linda Nøstbakken**, "The Cost of Endangered Species Protection: Evidence from Auctions for Natural Resources"

07/16 April, Cheti Nicoletti, **Kjell G. Salvanes**, and Emma Tominey, "The Family Peer Effect on Mothers' Labour Supply"

08/16 April, Hugh Gravelle and **Fred Schroyen**, "Optimal hospital payment rules under rationing by random waiting"

09/16 May, Branko Bošković and **Linda Nøstbakken** "Do land markets anticipate regulatory change? Evidence from Canadian conservation policy"

10/16 May, **Ingvild Almås**, Eleonora Freddi, and **Øystein Thøgersen**, "Saving and Bequest in China: An Analysis of Intergenerational Exchange"

11/16 May, **Itziar Lazkano, Linda Nøstbakken**, and Martino Pelli, "From Fossil Fuels to Renewables: The Role of Electricity Storage"

12/16 June, Jari Ojala and **Stig Tenold**, «Maritime trade and merchant shipping: The shipping/trade-ratio from the 1870s until today"

13/16 September, **Elias Braunfels**, "Further Unbundling Institutions"

14/16 September/April 2017, **Alexander W. Cappelen, Cornelius Cappelen, and Bertil Tungodden**», "False positives and false negatives in income distribution"

15/16 October, **Aline Bütikofer**, Eirin Mølland, and **Kjell G. Salvanes**, "Childhood Nutrition and Labor Market Outcomes: Evidence from a School Breakfast Program"

16/16 October**, Itziar Lazkano** and Linh Pham, "Do Fossil-Fuel Taxes Promote Innovation in Renewable Electricity Generation?

17/16 October, Kristiina Huttunen, Jarle Møen, and **Kjell G. Salvanes**, «Job Loss and Regional Mobility».

18/16 November**, Ingvild Almås, Alexander Cappelen, Bertil Tungodden**, «Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?"

19/16 December, **Gernot Doppelhofer, Ole-Petter Moe Hansen**, Melwyn Weeks, "Determinants of long-term economic growth redux: A Measurement Error Model Averaging (MEMA) approach "

20/16 December, Leroy Andersland and **Øivind A. Nilsen**, "Households' responses to price changes of formal childcare"

21/16 December, Wilko Letterie and **Øivind A. Nilsen**, "Price Changes - Stickiness and Internal Coordination in Multiproduct Firms"

22/16 December, **Øivind A. Nilsen** and Magne Vange, "Intermittent Price Changes in Production Plants: Empirical Evidence using Monthly Data"

2017

01/17 January, **Agnar Sandmo**, "Should the marginal tax rate be negative? Ragnar Frisch on the socially optimal amount of work"

02/17 February, **Luca Picariello**, "Organizational Design with Portable Skills"

03/17 March, **Kurt R. Brekke**, Tor Helge Holmås, Karin Monstad og Odd Rune Straume, "Competition and physician behaviour: Does the competitive environment affect the propensity to issue sickness certificates?"

04/17 March, **Mathias Ekström**, "Seasonal Social Preferences".

05/17 April, Orazio Attanasio, Agnes Kovacs, and **Krisztina Molnar**: "Euler Equations, Subjective Expectations and Income Shocks"

06/17 April, **Alexander W. Cappelen**, Karl Ove Moene, Siv-Elisabeth Skjelbred, and **Bertil Tungodden**, "The merit primacy effect"

07/17 May, **Jan Tore Klovland**, "Navigating through torpedo attacks and enemy raiders: Merchant shipping and freight rates during World War I"

08/17 May, **Alexander W. Cappelen**, Gary Charness, **Mathias Ekström**, Uri Gneezy, and **Bertil Tungodden: "**Exercise Improves Academic Performance"

09/17 June, **Astrid Kunze**, "The gender wage gap in developed countries"

10/17 June, **Kristina M. Bott, Alexander W. Cappelen, Erik Ø. Sørensen** and **Bertil Tungodden**, "You've got mail: A randomized field experiment on tax evasion"

11/17 August, Marco Pagano and **Luca Picariello**, "Talent Discovery, Layoff Risk and Unemployment Insurance"

12/17 August, Ingrid Kristine Folgerø, **Torfinn Harding** and Benjamin S. Westby, «Going fast or going green? Evidence from environmental speed limits in Norway"

13/17 August, **Chang-Koo Chi**, Pauli Murto, and Juuso Välimäki, "All-pay auctions with affiliated values"

14/17 August, **Helge Sandvig Thorsen**, "The effect of school consolidation on student achievement".

15/17 September, Arild Sæther, "Samuel Pufendorf and Ludvig Holberg on Political Economy".

16/17 September**, Chang-Koo Chi,** Pauli Murto, and Juuso Välimäki**, "**War of attrition with affiliated values".

17/17 September, **Aline Bütikofer** and Giovanni Peri, "The Effects of Cognitive and Noncognitive Skills on Migration Decisions"

18/17 October, **Øivind Schøyen**, "What limits the powerful in imposing the morality of their authority?

Norges Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no

Norges Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no