Supplementary Appendix for "The Construction of National Identities"

Milena Almagro
David Andrés-Cerezo

October 17, 2019

B Proof of Lemma 1

Lemma 1 *Under imperfect empathy and quadratic costs*

$$C(e) = \frac{1}{2}e^2,$$

the optimal socialization efforts are given by

$$e_t^N = (1-q_t)(1-\delta_t)r \qquad\qquad e_t^R = q_t\delta_t r,$$

and the law of motion for cultural transmission is

$$\dot{q} = q_t(1-q_t)\big(e_t^N - e_t^R\big) = r q_t(1-q_t)\big(1-\delta_t - q_t\big).$$

Proof.

A parent of trait i obtains utility V^{ij} if her child holds identity j. The imperfect empathy assumption implies parents evaluate children's actions using their own utility function. We assume a children of type i derives utility from private consumption but only consumes the public good associated to her identity (as the other provides zero utility). Therefore, for each combination of $i, j \in \{N, R\}$ one has

$$\begin{aligned} V^{NN} &= f(1-r) + (1-\delta_t)r & \qquad V^{RR} &= f(1-r) + \delta_t r \\ V^{NR} &= f(1-r) & \qquad V^{RN} &= f(1-r). \end{aligned} \tag{36}$$

Therefore, parents do not derive any utility from seeing their children consuming the club public good associated with the other identity.

Parents socialization problem for a type i parent is given by

$$\max_{e\in[0,1]} P_t^{ii}(e)V^{ii} + (1-P_t^{ii}(e))V^{ij} - \frac{1}{2}e^2, \tag{37}$$

with optimal socialization efforts

$$\begin{aligned} e_t^N &= (1-q_t)\Delta V^N = (1-q_t)g_t^N = (1-q_t)(1-\delta_t)r \\ e_t^R &= q_t \Delta V^R = q_t g_t^R = q_t \delta_t r, \end{aligned} \tag{38}$$

where $\Delta V^i = V^{ii} - V^{ij}$. Observe that for a parent of type i the optimal socialization effort depends positively on ΔV^i and negatively on q^i. The term ΔV^i is the degree of *cultural intolerance*, which increases in the level of own identity public good. Next we construct the evolutionary dynamics of cultural traits. Between t and $t+dt$, a fraction dt of the population dies and is replaced by the same number of new agents. Hence, at each point in time, type N proportion is given by the remaining parents of type N plus the fraction of newly born children inheriting trait N. Therefore, the fraction of agents with a national trait at time $t+dt$, q_{t+dt}, is

$$q_{t+dt} = (1-dt)q_t + dt\big[q_t P^{NN} + (1-q_t)P^{RN}\big]. \tag{39}$$

Recall that transition probabilities P^{NN} and P^{RN} are given by

$$P_t^{NN}(e_t^N) = e_t^N + (1-e_t^N)q_t \qquad\qquad P_t^{RN}(e_t^R) = (1-e_t^R)q_t \tag{40}$$

Using 38, 39 and 40, and taking $dt \to 0$, we obtain the following differential equation for q_t

$$\dot{q} = q_t(1-q_t)\big(e_t^N - e_t^R\big) = q_t(1-q_t)\big(1-\delta_t - q_t\big)r. \tag{41}$$

■

C Micro-foundations for the rates of protests

In this section we provide microfoundations for the individual decision on whether to participate in protests and we present different alternatives on how protests affect the objective function of the government.

C.1 Participation rate in protests

We rely on a stylized version of the model of political unrest developed by Passarelli and Tabellini (2017). As in their model, we assume that individuals engage in political unrest if the benefits of participating are greater than the costs. We also assume that the benefits of protesting are purely emotional rewards. That is, individuals join protests due to feelings of aggrievement and to the psychological reward that participating in protests provides to the individual. Following Passarelli and Tabellini (ibid.), we assume that individuals with identity i feel entitled to a particular policy $\widehat{g^i(\delta_t)}$. If this "reference" point is not implemented, individuals experiment a sense of injustice that causes them anger and frustration. The psychological reward of joining others in a protest is concomitant to this feeling of being treated unfairly. The further away actual policy is from their ideal point of a group of citizens, the more aggrieved they feel and the more they enjoy protesting.[1]

Formally, the emotional benefit of protesting $B^i(\cdot)$ is a function of the distance between their ideal policy $\widehat{g^i(\delta_t)}$ and the actual policy $g^i(\delta_t)$. In principle, emotional benefits could also depend on how many members from the group participate. Therefore, individual benefits from protesting are given by

$$B^i\Big(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t^i\Big) = F\Big(\text{dist}\Big(\widehat{g^i(\delta_t)}, g^i(\delta_t)\Big), q_t^i\Big) = \text{dist}\Big(\widehat{g^i(\delta_t)}, g^i(\delta_t)\Big) \times h(q_t^i)$$

with dist defined as some distance, and $h(\cdot)$ an arbitrary function to be defined later. This specification allows for several specifications depending on the choice of $dist(\cdot)$, $h(\cdot)$ and $\widehat{g^i(\delta_t)}$.

However, joining protests is costly. Concretely, we assume that individuals in group i face cost c, independently drawn from some distribution F. These costs capture common features such as repression as well as idiosyncratic costs, such as foregone income from not working. Thus, individual j in group i participates in protests if and only if $B^i\Big(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t\Big) - c^{ij} \geq 0$. Hence, if $c^{ij} \sim U[0,1]$, the individual probability of engaging in protests is given by

$$p_t^i = Pr\Big(c^{ij} \leq B^i(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t)\Big) = B^i(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t)$$

Therefore, the total participation rate $P^i(\delta_t, q_t)$ in protests of group i is given by

$$D^i(\delta_t, q_t) = q_t^i \times p_t^i = q_t^i \times B^i(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t)$$

Finally, as we discuss below, protests affect the objective function of the central government, either by creating a direct welfare loss for the government, or indirectly by generating dead-weight losses for citizens which in turn are internalized by a welfarist government.

[1]One could argue that the choice to participate in a riot or a civil conflict should based on individual expectations about how joining a protest changes the policy choices of the central government. Although we recognize that this "instrumental" motive has its merits, we believe that it is not very relevant in our context. In a sufficiently large and heterogeneous population of potential protesters, which is generally the case in our context, the marginal impact of one more individual protesting in the decision of the government is negligible. Hence, an atomistic individual is unlikely to take this costly political action. Given that the expected change in welfare through influencing policy choices is close to zero, and in the absence of any explicit material gain of protesting, the benefit from protesting must come from psychological or social rewards. In our case, as argued by Laitin (2007), a key feature of national identities is the willingness that creates on individuals to engage in costly political actions, in order to defend their own nation for the psychological reward that provides and despite obvious material losses.

C.1.1 Benchmark case

In the benchmark case, we assume the following

- $dist(x,y) = |x-y|$
- $h(q_t^i) = 1$
- $\widehat{g^N(\delta_t)} = r$ and $\widehat{g^R(\delta_t)} = r$

That is, the benefits of protesting depend linearly on the distance between the ideal policy and the policy implemented, and individual emotional rewards are orthogonal to the number of individuals participating.[2] Also, we assume an extreme polarization of preferences, in the sense that members of each group feel entitled to a level of public good equal to the total tax collection in the region i.e. the ideal δ^i for each group is $\widehat{\delta_t} = 0$ for type N and $\widehat{\delta_t} = 1$ for type R. Therefore, we have that

$$D_t^N(\delta_t, q_t) = q_t[\underbrace{r}_{\text{Ideal}} - \underbrace{(1-\delta_t)r}_{\text{Real}}] = q_t \delta_t r$$
$$D_t^R(\delta_t, q_t) = (1-q_t)[\underbrace{r}_{\text{Ideal}} - \underbrace{\delta_t r}_{\text{Real}}] = (1-q_t)(1-\delta_t)r$$

Finally, for the baseline case we assume that the government directly experiments a loss of welfare which is proportional to the participation in protests of both groups. Therefore, the utility function of the central government is

$$\begin{aligned} W(q_t, \delta_t) = \psi^N q_t + \alpha q_t U^N(\delta_t) + (1-\alpha)(1-q_t)U^R(\delta_t) \\ -\big(\beta q_t \delta_t r + (1-\beta)(1-q_t)(1-\delta_t)r\big), \end{aligned}$$

where β and $1-\beta$ capture the disruptions created by protests, which inflict a direct loss of social welfare to the central government. In this setting, β is a measure for the relative impact of protests of group N with respect to group R, and it comprises factors such as how organized individuals are, the capacity of regional cultural leaders to mobilize people along identity cleavages, the physical resources they have to cause disruption, their influence on media or the support they have from international public opinion.[3]

[2] In all the specifications of the protest function we assume that the individual decision about participating in protests is independent of the number of members from her group joining the protest i.e. $h(q_t^i) = 1$. An interesting possibility is to allow for complementarities in protests. Concretely, we could assume that the individual emotional benefit increases with the number of individuals that also participate in protests i.e. $h(q_t^i) = p_t^i q_t^i$, where p_t^i is the average participation rate at time t of individuals in group i. Nevertheless, the main qualitative results of the paper are robust to these type of protests.

[3] In the context of the paper, D_t^N and D_t^R can also captures the idea that political unrest above some threshold could generate violent civil conflict and a secessionist attempt in the peripheral region. Then, the participation rate can be interpreted as the probability of reaching that turning point.

C.1.2 Quadratic case

In section 5 we illustrate how the results of the model change when we relax the linearity assumption of the objective function. Concretely, we keep the rest of the assumptions but we have that $dist(x,y) = (x-y)^2$ instead of $dist(x,y) = |x-y|$. Therefore, protests are given by

$$D_t^N(\delta_t, q_t) = q_t[\underbrace{r}_{\text{Ideal}} - \underbrace{(1-\delta_t)r}_{\text{Real}}]^2 = q_t \delta_t^2 r^2$$
$$D_t^R(\delta_t, q_t) = (1-q_t)[\underbrace{r}_{\text{Ideal}} - \underbrace{\delta_t r}_{\text{Real}}]^2 = (1-q_t)(1-\delta_t)^2 r^2$$

C.1.3 Different ideal point

The previous choice of the ideal point, which is a maintained assumption throughout the paper, corresponds to a very extreme case in which individuals in both groups are entirely selfish. However, considering ideal points that involve some sharing of resources may be more reasonable for some real-world examples.[4] Moreover, it may be that this assumption is behind the full homogenization result, as it introduces a strong conflict over resources. Nevertheless, it turns out that our homogeneity results are robust to ideal points that incorporate some fairness concerns.

To see this, consider that protests have the same structure as in the benchmark model but ideal points are defined as follows

$$\widehat{g^i(\delta_t)} = r(\lambda^N + (1-\lambda^N)q_t)$$
$$\widehat{g^i(\delta_t)} = r(\lambda^R + (1-\lambda^R)(1-q_t))$$

where a higher $\lambda^i \in [0,1]$ implies a higher degree of selfishness of individuals in group i. Note that

$$\lim_{\lambda^i \to 1} \widehat{g^i(\delta_t)} = r$$
$$\lim_{\lambda^i \to 0} \widehat{g^i(\delta_t)} = q_t^i r$$

Therefore, the formulation of ideal points has two extreme cases: 1) the one in the paper, where citizens are entirely selfish; 2) the "perfectly fair" case, where individuals feel entitled to get in public goods a fraction of the budget equal to the size of their

[4]We thank the editor for his suggestion about checking the robustness of the results to less extreme choices of ideal points.

group in the population. The value of λ^i captures the self-serving bias of the individuals in group i, as individuals judgments combine what is fair and what is beneficial for them.

Now, consider a situation where individuals protest whenever the policy deviates from their bliss point, even if it is beneficial to them. For comparability with results in Section 5, also consider quadratic protests. The protest functions are given by

$$D_t^N(\delta_t, q_t) = q_t\big[\widehat{g^N(\delta_t)} - g^N(\delta_t)\big]^2 = q_t\big[\underbrace{r(\lambda^N + (1-\lambda^N)q_t)}_{\text{Ideal}} - \underbrace{(1-\delta_t)r}_{\text{Real}}\big]^2$$
$$D_t^R(\delta_t, q_t) = (1-q_t)\big[[\widehat{g^R(\delta_t)} - g^R(\delta_t)\big]^2 = (1-q_t)\big[\underbrace{r(\lambda^R + (1-\lambda^R)(1-q_t))}_{\text{Ideal}} - \underbrace{\delta_t r}_{\text{Real}}\big]^2$$

From now on, we assume that $\lambda^i = \lambda, \forall i$, as it simplifies the algebra (but the results below hold for any combination of λ^N and λ^R).

If the function $H(q)$ for this problem is strictly convex for all q, then Theorem 3 holds, so long run steady states are homogeneous. Recall that $H(q)$ gives the per-period utility derived from the policy $\delta(q)$ that keeps q unchanged. We have that

$$H(q) = \psi^N q + \big(\alpha q + (1-\alpha)(1-q)\big)f(1-r) + r\big(\alpha q^2 + (1-\alpha)(1-q)^2\big)$$
$$-r^2\Big[\beta q_t\big[(\lambda + (1-\lambda)q_t) - q_t\big]^2 + (1-\beta)(1-q_t)\big[(\lambda + (1-\lambda)(1-q_t)) - (1-q_t)\big]^2.$$

The second derivative of this function is given by

$$H''(q) = 2r + 2r^2\lambda^2\big[\beta(1-q) + (1-\beta)q - (1-2q)(1-2\beta)\big].$$

Observe for all $\beta, q, \lambda, r \in [0,1]$

$$\beta(1-q) + (1-\beta)q - (1-2q)(1-2\beta) \geq -1.$$

Hence,

$$H''(q) \geq 2r - 2r^2\lambda^2 = 2r(1 - r\lambda^2) \geq 0.$$

Therefore, $H(q)$ does not have a maximum in $[0,1]$ for any choice of ideal point. Hence, long-run steady states are culturally homogeneous.[5]

[5]Another possibility is to assume that individuals only protest when the deviation is detrimental for them. In this case, they may do nothing (zero protests) or they may show support for the government if it benefits them ("positive" protests). That is

$$D_t^i(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t^i) = q_t^i \max\{\widehat{g^i(\delta_t)} - g^i(\delta_t), 0\}$$

or

$$D_t^i(\widehat{g^i(\delta_t)}, g^i(\delta_t), q_t^i) = q_t^i\big[\widehat{g^i(\delta_t)} - g^i(\delta_t)\big]$$

Although we do not present it here, the same result goes through if we consider these alternative for-

In conclusion, allowing for ideal points that involve some sharing of resources does not alter the full-homogenization result. When the two groups have closer views about what they are entitled to (lower λ), the zero-sum conflict is weakened because the government can reduce the utility losses coming from protests by choosing a value of δ close to the ideal point of both groups. However, the conflict never completely disappears as long as there is a heterogeneous distribution of identities. The reason is that it is unavoidable for the government to pick winners and losers, as a larger provision of one public good always comes at the expense of a reduction in the other public good. Therefore, the government can only avoid dealing with conflicting motives by homogenizing the population.

C.2 Alternative rationales for the objective function

One could think of alternative rationales for how protests affect the objective function of the government. One possibility is to assume that citizens experiment a direct intrinsic utility loss from seeing the other group protesting, which in turn is internalized by the government, as it cares about the utilities of individuals. In the same way as protesting to defend one's identity provides an emotional reward (by singing the anthem, carrying the flag, etc...), seeing protests by the group with the oppositional identity can create feelings of anger and reductions of self and group-esteem. Let β and $1-\beta$ be the marginal disruption created by protests of groups N and R, respectively. Then we can write

$$\begin{aligned} U^N(\cdot) &= f((1-r)) + (1-\delta_t)r - (1-\beta)D^R(\delta_t, q_t) \\ U^R(\cdot) &= f((1-r)) + \delta_t r - \beta D^N(\delta_t, q_t) \end{aligned}$$

Therefore, we have

$$\begin{aligned} W &= \psi^N q_t + \alpha q_t U_t^N(\cdot) + (1-\alpha)(1-q_t)U_t^R(\cdot) \\ &= \psi^N q_t + \alpha q_t\big[f((1-r)) + (1-\delta_t)r - (1-\beta)D^R(\delta_t, q_t)\big] \\ &\quad + (1-\alpha)(1-q_t)\big[f((1-r)) + \delta_t r - \beta D^N(\delta_t, q_t)\big] \end{aligned}$$

We can see that this objective function is similar to the previous one, with a higher order term for q on the protest side. In this case, protests of both groups are higher at intermediate values of q, which makes homogeneous steady-states more desirable.

Another alternative is to assume that, in order to keep order and counteract the disruptive costs of protests, the government uses revenue from taxes, which is taken away from the total public budget used to provide public goods. To keep comparability, we can assume that in order to repair the damage created by protests, the government needs to employ a fraction ζ and a fraction η of the public budget r to counteract

mulations of the protest function, *for any choice of ideal point.*

protest by N and R respectively. Therefore, $g_t^N + g_t^R = r\big[1 - \zeta D^N(q_t, \delta_t) - \eta D^R(q_t, \delta_t)\big]$.[6] We can also assume that, in addition to the destruction of public goods, riots have an effect on disposable (after tax) income. For instance, this would be due to the shutdown of economic activity, the increase in risk premium of bonds or the destruction of physical capital needed to generate income. In both cases, we will get a very similar objective function.

These different rationales tell slightly different plausible stories about the processes of nation-building. However, the different models are formally equivalent and their qualitative results identical. In some sense, one can move from one to another by relabelling parameters, as the key results are robust to the chosen specification.

D Proofs of Propositions 2, 3, and 4

D.1 Proof of Proposition 2

First we prove the difference in welfare of the two policies is decreasing at $q_0 = \bar{q}_0$; that is

$$\frac{\partial}{\partial q} F(\bar{q}_0) < 0.$$

First, we rule out $\frac{\partial}{\partial q} F(\bar{q}_0) = 0$. Simply observe for any variable x

$$F(\bar{q}_0(x), x) \equiv 0$$

Hence, for any variable x, it follows

$$\frac{\partial}{\partial q} F(\bar{q}_0(x), x) \frac{\partial}{\partial x} \bar{q}_0(x) + \frac{\partial}{\partial x} F(\bar{q}_0(x), x) = 0.$$

It is easy to verify $\frac{\partial}{\partial q} F(\bar{q}_0(x), x) < 0$. Assume for a contradiction $\frac{\partial}{\partial q} F(\bar{q}_0) \geq 0$. Then, because F is continuous in q_0 and $F(\bar{q}_0) = 0$ it must be the case that $\exists q_0' > \bar{q}_0$ such that $F(q_0') \geq 0$. Because F is continuous with $F(1) < 0$, we contradict the result that $F(.)$ has a unique zero.

Now because $\frac{\partial}{\partial q} F(\bar{q}_0(x), x) \neq 0$, we can write

$$\frac{\partial}{\partial x} \bar{q}_0 = -\left(\frac{\partial}{\partial q} F(\bar{q}_0)\right)^{-1} \frac{\partial}{\partial x} F(\bar{q}_0).$$

But since $\frac{\partial}{\partial q} F(\bar{q}_0) < 0$, we have that

$$sign\Big(\frac{\partial}{\partial x} \bar{q}_0(\omega)\Big) = sign\Big(\frac{\partial}{\partial x} F(\bar{q}_0(\omega); \omega)\Big).$$

[6] We need to assume that η and ζ are sufficiently small so that $g_t^N + g_t^R \geq 0$

Hence, for parameter x, we only need to check the sign of

$$\frac{\partial}{\partial x} F(\bar{q}_0(x), x).$$

Recall that

$$\begin{aligned}W^{\delta}(q;\omega) = \psi^N q + \\ &+ \alpha q\big(f(1-r)+(1-\delta)r\big) \\ &+(1-\alpha)(1-q)\big(f(1-r)+\delta r\big) \\ &-\beta q \delta r \\ &-(1-\beta)(1-q)(1-\delta)r,\end{aligned}$$

where ω is a vector including all the parameters. Let's denote

$$\begin{aligned}S(q^1, q^0, \omega) &= W^1(q^1(q_0;\omega);\omega) - W^0(q^0(q_0;\omega);\omega) \\ &= \psi^N(q^1-q^0) + (2\alpha-1)f(1-r)(q_1-q_0) + r\big((1-\alpha)(1-q_1)-\alpha q_0\big) \\ &+ r\big((1-\beta)(1-q_0)-\beta q_1\big)\end{aligned}$$

It follows

$$\frac{\partial}{\partial x} F(q_0;\omega) = \int_0^{\infty} \frac{\partial}{\partial x} e^{-\rho t} S(q_t^1(\omega), q_t^0(\omega), \omega) dt,$$

for any parameter x. Next, we do comparative statics on the parameters of the model.

- We begin with the comparative statics for ψ^N. These are as follows

$$\frac{\partial}{\partial \psi^N} F(\bar{q}_0) = \int_0^{\infty} e^{-\rho t}(q_t^1 - q_t^0) dt < 0,$$

  because we always have that $q_t^0 > \bar{q}_0 > q_t^1$ for all $t > 0$. Therefore

$$\frac{\partial \bar{q}_0}{\partial \psi^N} < 0.$$

- Next, we do comparative statics for α. These are as follows

$$\frac{\partial}{\partial \alpha} F(\bar{q}_0) = \int_0^{\infty} e^{-\rho t}\left\{2f(1-r)(q_t^1 - q_t^0) - r\big(q_t^0 + (1-q_t^1)\big)\right\} dt < 0,$$

  hence

$$\frac{\partial \bar{q}_0}{\partial \alpha} < 0.$$

- Clearly, for parameter β, we obtain similar results

$$\frac{\partial}{\partial \beta} F(\bar{q}_0) = \int_0^{\infty} e^{-\rho t}\left\{-r\big((1-q_t^0) + q_t^1\big)\right\} dt < 0,$$

Therefore

$$\frac{\partial \bar{q}_0}{\partial \beta} < 0.$$

- Now if utility of consumption is given by $f(1-r) = \frac{(1-r)^{1-\sigma}}{1-\sigma}$, where $\theta \in (0,1), \sigma > 0$, it holds

$$\frac{\partial}{\partial \theta} F(\bar{q}_0) = \int_0^\infty e^{-\rho t}\{(2\alpha-1)\frac{(1-r)^{1-\sigma}}{1-\sigma}(q_t^1 - q_t^0)\} dt > 0 \iff \alpha < \frac{1}{2}$$

$$\frac{\partial}{\partial \sigma} F(\bar{q}_0) = \int_0^\infty e^{-\rho t}\Big\{(2\alpha-1)\theta\frac{(1-r)^{1-\sigma}}{(1-\sigma)^2}\big(1-\ln(1-r)(1-\sigma)\big)(q_t^1 - q_t^0)\Big\} dt > 0$$

$$\iff \alpha < \frac{1}{2}$$

Hence, it follows

$$\frac{\partial \bar{q}_0}{\partial \theta} > 0 \iff \alpha < \frac{1}{2},$$
$$\frac{\partial \bar{q}_0}{\partial \sigma} > 0 \iff \alpha < \frac{1}{2},$$

because $q_t^1 - q_t^0 < 0$, $(1-r)^{1-\sigma} > 0$, and $1-\ln(1-r)(1-\sigma) > 0$ for all $r \in (0,1)$.

Now we show how the comparative statics on ρ and r can go both ways.

D.2 Proof of Proposition 3

Next we do comparative statics on ρ. Taking derivatives of $F(q_0)$ with respect to ρ, we obtain the following expression

$$\frac{\partial}{\partial \rho} F(q_0) = \int_0^\infty \frac{\partial}{\partial \rho} e^{-\rho t} S(q_t^1, q_t^0) dt =$$
$$-\int_0^\infty t e^{-\rho t} S(q_t^1, q_t^0) dt. \tag{42}$$

It is easy to see $S(q^1, q^0)$ is bounded. Hence, an $M > 0$ exists such that $|S(q_t^1, q_t^0)| \leq M$. For example, we can pick $M = f(1-r)$ whenever $f(c) = \frac{x^{1-\sigma}}{1-\sigma}$, $\gamma \geq 0$. Therefore,

$$\left|\frac{\partial}{\partial \rho} F(q_0;\omega)\right| \leq \int_0^\infty t e^{-\rho t} |S(q_t^1(\omega), q_t^0(\omega), \omega)| dt < \int_0^\infty t e^{-\rho t} M dt = \frac{1}{\rho^2} M < \infty,$$

and the integral 42 is always well-defined.

Recall that the function $S(q^1, q^0)$ can be written as

$$S(q^1, q^0) = (A_1 q^1 + B_1) - (A_0 q_0 + B_0),$$

with

$$A_1 = \psi^N + (2\alpha - 1) f(1-r) - (1-\alpha+\beta) r, \qquad B_1 = (1-\alpha)(f(1-r)+r)$$
$$A_0 = \psi^N + (2\alpha - 1) f(1-r) + (\alpha+1-\beta) r, \qquad B_0 = (1-\alpha) f(1-r) - (1-\beta) r$$

The sign of the comparative statics on ρ can go both ways as it will depend on the other parameters of the model. Hence, we analyze different cases.

- Assume, α large enough such that $A_1 > 0$, which implies $A_0 > 0$. It is easy to see that
$$S(q_t^1, q_t^0) = A_1 q_t^1 + B_1 - (A_0 q_t^0 + B_0),$$
is strictly decreasing in t with
$$\begin{aligned}\lim_{t\to\infty} S(q_t^1, q_t^0) &= B_1 - A_0 - B_0 = -\psi^N + (1-2\alpha)\big(f(1-r)+r\big)\\ &= -(A_1 + (\alpha+\beta) r) < 0\end{aligned}$$
Because $\int_0^\infty e^{-\rho t} S(q_t^1, q_t^0) S(q_t^1, q_t^0) dt = 0$ with $S(q_t^1, q_t^0)$ is strictly decreasing, and
$$\lim_{t\to\infty} S(q_t^1, q_t^0) < 0,$$
a T exists such that $S(q_T^1, q_T^0) = 0$, with $S(q_t^1, q_t^0) > 0$ for all $t \le T$, and $S(q_t^1, q_t^0) < 0$ for all $t \ge T$. It follows
$$\begin{aligned}\frac{\partial}{\partial \rho} F(\bar{q}_0) &= -\int_0^\infty t e^{-\rho t} S(q_t^1, q_t^0) dt\\ &= -\int_0^T t e^{-\rho t} S(q_t^1, q_t^0) dt - \int_T^\infty t e^{-\rho t} S(q_t^1, q_t^0) dt\\ &> -\int_0^T T e^{-\rho t} S(q_t^1, q_t^0) dt - \int_T^\infty \big((t-T)+T\big) e^{-\rho t} S(q_t^1, q_t^0) dt\\ &= -T \int_0^\infty e^{-\rho t} S(q_t^1, q_t^0) dt - \int_T^\infty \big(t-T\big) e^{-\rho t} S(q_t^1, q_t^0) dt\\ &> 0 - \int_T^\infty \big(t-T\big) e^{-\rho t} S(q_t^1, q_t^0) dt > 0,\end{aligned}$$
because $t - T \ge 0$ for all $t \ge T$ and $S(q_t^1, q_t^0) < 0$ for all $t > T$. The last inequality implies
$$\frac{\partial}{\partial \rho} \bar{q}_0 > 0.$$
- Assume α small enough such that $A_0 < 0$ which implies $A_1 < 0$. Following a simi-

lar argument, we obtain

$$\frac{\partial}{\partial \rho}F(\bar{q}_0) = -\int_0^{\infty} t e^{-\rho t} S(q_t^1, q_t^0) dt < 0,$$

implying in turn

$$\frac{\partial}{\partial \rho}\bar{q}_0 < 0.$$

In this way, we have shown that the comparative statics on ρ can go in both directions. The following result summarizes the previous discussion

Proposition 19 *The comparative statics on ρ can go both ways and depend on the other parameters of the model*

- *If* $\psi^N + (2\alpha - 1)f(1-r) - r(1-\alpha+\beta) > 0$*, then*
$$\frac{\partial}{\partial \rho}\bar{q}_0 > 0.$$
- *If* $\psi^N + (2\alpha - 1)f(1-r) + r(\alpha+1-\beta) < 0$*, then*
$$\frac{\partial}{\partial \rho}\bar{q}_0 < 0.$$

D.3 Proof of Proposition 4

Finally, we do comparative statics on r.

Proposition 20 *The following equality holds*

$$\frac{\partial}{\partial r}F(\bar{q}_0) = -\frac{\rho}{r}\frac{\partial}{\partial \rho}F(\bar{q}_0) + \Lambda \int_0^{\infty} e^{-\rho t}(q_t^0 - q_t^1) dt \tag{43}$$

with $\Lambda = (2\alpha-1)f'(1-r) + \frac{1}{r}\big((2\alpha-1)f(1-r) + \psi^N\big)$.

Proof. Given that r enters the low of motion we have

$$\frac{d}{dr}S(q^1, q^0) = \frac{\partial}{\partial q^1}S\frac{\partial}{\partial r}q^1 + \frac{\partial}{\partial q^0}S\frac{\partial}{\partial r}q^0 + \frac{\partial}{\partial r}S, \tag{44}$$

where the first two terms come from r entering in the law of motion and the third terms comes from r entering in the function S. First, observe we can write

$$G(q_t) = rt + G(q_0)$$

where $G'(y) = \frac{1}{g(y)}$ with $\dot{q} = rg(q)$. Therefore, taking derivatives with respect to r on both sides of the previous expression

$$\frac{\partial}{\partial r}q_t = \frac{t}{G'(q_t)} = t g(q_t) = \frac{t}{r}\dot{q}_t,$$

Finally observe

$$
\begin{aligned}
\int_0^\infty e^{-\rho t}\frac{t}{r}\dot q_t\,dt &= \Big[\frac{t}{r}e^{-\rho t}q_t\Big]_0^\infty - \int_0^\infty \frac{1}{r}e^{-\rho t}(1-\rho t)q_t\,dt \\
&= -\int_0^\infty \frac{1}{r}e^{-\rho t}(1-\rho t)q_t\,dt,
\end{aligned}
$$

where we have used integration by parts.

Recall that the function $S(q^1,q^0)$ can be written as

$$S(q^1,q^0)=(A_1q^1+B_1)-(A_0q^0+B_0),$$

with

$$
\begin{aligned}
A_1 &= \psi^N+(2\alpha-1)f(1-r)-(1-\alpha+\beta)r, && B_1=(1-\alpha)(f(1-r)+r),\\
A_0 &= \psi^N+(2\alpha-1)f(1-r)+(\alpha+1-\beta)r, && B_0=(1-\alpha)f(1-r)-(1-\beta)r,
\end{aligned}
$$

so the integral of the first two terms of expression 44 are given by

$$
\begin{aligned}
&\int_0^\infty e^{-\rho t}\Big(\frac{\partial}{\partial q^1}S\frac{\partial}{\partial r}q^1+\frac{\partial}{\partial q^0}S\frac{\partial}{\partial r}q^0\Big)dt = \int_0^\infty e^{-\rho t}\Big(A_1\frac{t}{r}\dot q_t^1-A_0\frac{t}{r}\dot q_t^0\Big)dt\\
&=-\int_0^\infty e^{-\rho t}\frac{1}{r}(1-\rho t)\big\{S(q_t^1,q_t^0)-(B_1-B_0)\big\}dt\\
&=\frac{\rho}{r}\int_0^\infty e^{-\rho t}\,t\,S(q_t^1,q_t^0)dt-\frac{1}{r}(B_1-B_0)\int_0^\infty e^{-\rho t}(1-\rho t)dt\\
&=\frac{\rho}{r}\int_0^\infty e^{-\rho t}\,t\,S(q_t^1,q_t^0)dt-r(B_1-B_0)\Big[t\,e^{-\rho t}\Big]_0^\infty\\
&=\frac{\rho}{r}\int_0^\infty e^{-\rho t}\,t\,S(q_t^1,q_t^0)dt && (45)\\
&=-\frac{\rho}{r}\frac{\partial}{\partial \rho}F(\bar q_0) && (46)
\end{aligned}
$$

The third term in expression 44 is given by

$$
\begin{aligned}
&\Big(-(2\alpha-1)f'(1-r)-(1-\alpha+\beta)\Big)q^1+(1-\alpha)(-f'(1-r)+1)\\
&-\Big(-(2\alpha-1)f'(1-r)+(\alpha+1-\beta)\Big)q^0+(1-\alpha)f'(1-r)+(1-\beta)=\\
&\Big((2\alpha-1)f'(1-r)+\frac{1}{r}\Big((2\alpha-1)f(1-r)+\psi^N\Big)\Big)(q^0-q^1)+\frac{1}{r}S(q^1,q^0). \qquad (47)
\end{aligned}
$$

Denote

$$\Lambda\equiv\Big((2\alpha-1)f'(1-r)+\frac{1}{r}\Big((2\alpha-1)f(1-r)+\psi^N\Big)\Big).$$

Integrating expression 47

$$\int_0^\infty e^{-\rho t}\frac{\partial}{\partial r}S(q_t^1,q_t^0)dt = \int_0^\infty e^{-\rho t}\Big\{\Lambda(q_t^0-q_t^1)+\frac{1}{r}S(q_t^1,q_t^0)\Big\}dt$$
$$= \Lambda\int_0^\infty e^{-\rho t}\Big(q_t^0-q_t^1\Big)dt. \tag{48}$$

Combining 46 and 48, we obtain

$$\frac{\partial}{\partial r}F(\bar{q}_0) = \int_0^\infty \frac{t}{r}\rho e^{-\rho t}S(q_t^1,q_t^0)dt + \Lambda\int_0^\infty e^{-\rho t}(q_t^0-q_t^1)dt$$
$$= -\frac{\rho}{r}\frac{\partial}{\partial\rho}F(\bar{q}_0)+\Lambda\int_0^\infty e^{-\rho t}(q_t^0-q_t^1)dt, \tag{49}$$

■

The first term of expression 49 captures the fact that ρ and r play opposite roles in our model: an increase in r makes dynamics faster, so it is effectively equal to moving any future point closer to the present, or equivalently, putting more weight into the future. Hence, an increase in r can be equivalently seen as a reduction in ρ. Besides the effect that r has on the dynamics, it also has an effect on individual utilities and protests, which is captured by the second term in 49.

Finally, using the last proposition we see that the comparative statics on r can also go both ways because they depend on the other parameters of the model.

Proposition 21 *The comparative statics on r can go both ways and depend on the other parameters of the model:*

- *For small α, and sufficiently large ψ^N, it follows*
$$\frac{\partial}{\partial r}\bar{q}_0 < 0.$$
- *On the other hand, for large α, and sufficiently small ψ^N, it follows*
$$\frac{\partial}{\partial r}\bar{q}_0 > 0.$$

Proof. Take small α and sufficiently large ψ^N such that

$$A_1 > 0 > \Lambda.$$

Using Proposition 20 we see

$$\frac{\partial}{\partial\rho}F(\bar{q}_0) > 0.$$

Combining the previous inequality with 20 and $\Lambda < 0$, we obtain

$$\frac{\partial}{\partial r}F(\bar{q}_0) = -\frac{\rho}{r}\frac{\partial}{\partial\rho}F(\bar{q}_0)+\int_0^\infty e^{-\rho t}(q_t^0-q_t^1)dt < 0,$$

which proves the first part of the proposition. The second part is proved similarly. ■

To complement our analysis, the following graphs show numerical solutions for the threshold $\bar{q}_0$. We fix the other parameters at $\alpha = 0.9, \beta = 0.5, \theta = 0.4, r = 0.3, \sigma = 0.5$, and $\rho = 0.5$ and let the corresponding parameter run over some range.

Baseline values $\psi^N = 0.5, \alpha = 0.9, \beta = 0.5, \theta = 0.3, \sigma = 0.5, r = 0.3, \rho = 0.5$

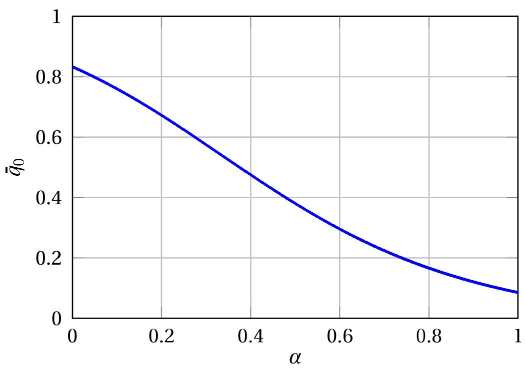

Figure 10: α

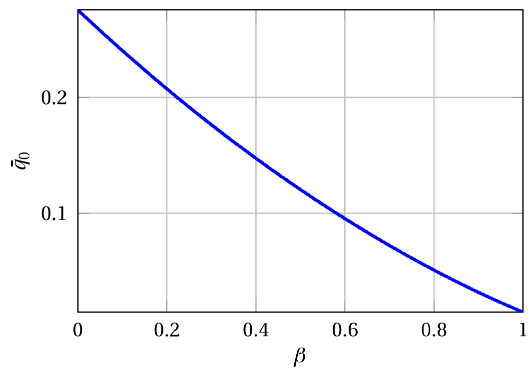

Figure 11: β

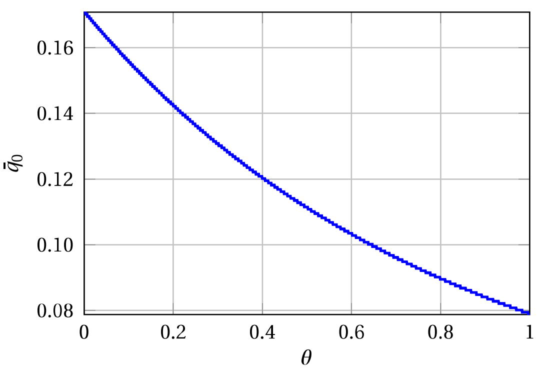

Figure 12: θ

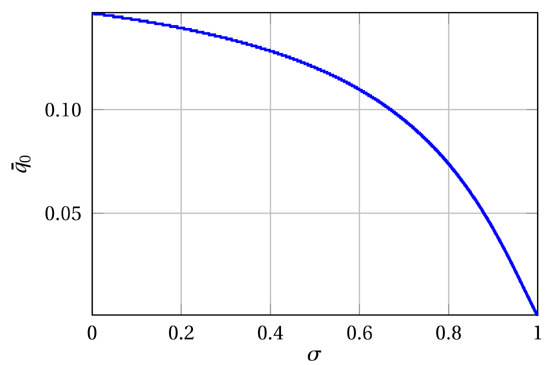

Figure 13: σ

For r and ρ, we show a case with $A_1 > 0$, where we choose $\alpha = 0.9, \beta = 0.5, \theta = 0.4, \sigma = 0.5, r = 0.3, \rho = 0.5$ as baseline parameters and plot the region in which the condition is satisfied.

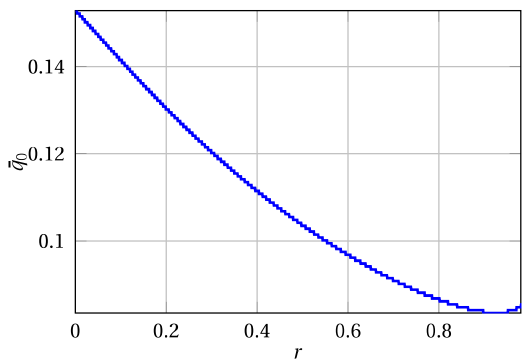

Figure 14: Case $A_1 > 0$, r

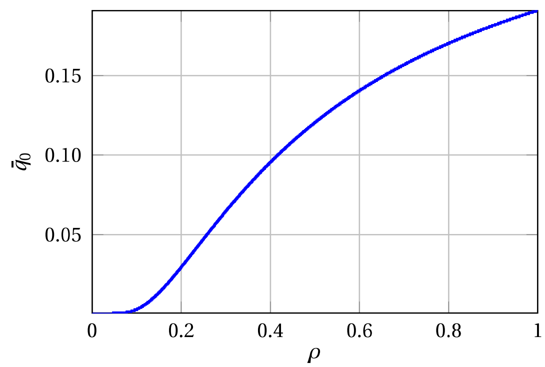

Figure 15: Case $A_1 > 0$, ρ

E Proofs and extra material for Section 5

E.1 Technical details for proof of Theorem 3

First, we prove $\frac{\partial}{\partial \tau}\big(W(q_t, \delta'_\tau(q_t)) - H(\tilde{q})\big) = 0$. Observe

$$\frac{\partial}{\partial \tau}\big(W(q_t, \delta'_\tau(q_t)) - H(\tilde{q})\big) = \frac{\partial}{\partial \delta} W(q_t, \delta'_\tau(q_t)) \frac{\partial}{\partial \tau} \delta'_\tau(q_t).$$

It holds

$$\delta'_\tau(q) = \mathbb{1}\{t < \tau\}(\delta^S(q) + \epsilon(q)) + \mathbb{1}\{t \geq \tau\}\delta^S(q),$$

therefore

$$\frac{\partial}{\partial \tau}\delta'(q) = \Delta(\tau)\big(\delta^S(q) + \epsilon(q)\big) - \Delta(\tau)\delta^S(q) = -\Delta(\tau)\epsilon(q),$$

where $\Delta(\tau)$ is the Dirac delta function

$$\Delta(\tau) = \begin{cases} 1 & \text{if } t = \tau \\ 0 & \text{if } t \neq \tau. \end{cases}$$

Integrating

$$\int_0^\tau \frac{\partial}{\partial \tau}\delta'_\tau(q) e^{-\rho t}(W(q_t, \delta'_\tau(q_t)) - H(\tilde{q}))dt = \int_0^\tau -\frac{\partial}{\partial \delta} W(q_t, \delta'_\tau(q_t))\Delta(\tau)\epsilon(q_t)dt = 0,$$

for all $\tau > 0$.

Second, we prove Theorem 3 still holds when $H'(\tilde{q}) = 0$ using a second order Taylor expansion for $F(\tau)$. We take derivatives with respect to τ from expression 27 to obtain an expression for $F''(\tau)$

$$\begin{aligned}F''(\tau) = &-\rho e^{-\rho\tau}\big(H(q_\tau) - H(\tilde{q})\big) + e^{-\rho\tau} H'(q_\tau)\dot{q}_\tau\\&+\rho^2 e^{-\rho\tau}\big(W(q_\tau, \delta'(q_\tau)) - H(\tilde{q})\big)\\&-\rho e^{-\rho\tau}\big(\frac{\partial}{\partial q} W(q_\tau, \delta'(q_\tau) + \frac{\partial}{\partial \delta} W(q_\tau, \delta'(q_\tau))\frac{\partial}{\partial q}\delta'(q_\tau)\big)\dot{q}_\tau\\&-\rho e^{-\rho\tau} H'(q_\tau)\dot{q}_\tau + e^{-\rho\tau} H''(q_\tau)(\dot{q}_\tau)^2.\end{aligned}$$

Observe $\frac{\partial}{\partial q} W(q_\tau, \delta'(q_\tau)) + \frac{\partial}{\partial \delta} W(q_\tau, \delta'(q_\tau))\frac{\partial}{\partial q}\delta'(q_\tau) = \frac{\partial}{\partial q} H(q_\tau) = H'(q_\tau)$, hence the previous expression simplifies to

$$\begin{aligned}F''(\tau) = &-\rho e^{-\rho\tau}\big(H(q_\tau) - H(\tilde{q})\big) + e^{-\rho\tau}(1-2\rho) H'(q_\tau)\dot{q}_\tau\\&+\rho^2 e^{-\rho\tau}\big(W(q_\tau, \delta'(q_\tau)) - H(\tilde{q})\big)\\&+ e^{-\rho\tau} H''(q_\tau)(\dot{q}_\tau)^2.\end{aligned}$$

Evaluating at $\tau = 0$

$$F''(0) = \frac{1}{\rho} H''(\tilde{q})(\dot{\tilde{q}})^2 > 0,$$

because $\tilde{q}$ is not local maximum of $H(q)$, with $H'(\tilde{q}) = 0$. Therefore, it must hold $H''(\tilde{q}) > 0$. Then

$$J(\tilde{q}, \delta'_\tau(\tilde{q})) - J(\tilde{q}, \delta^*(\tilde{q})) = F(\tau) > 0,$$

a contradiction.

E.2 Quadratic Protests

When protests enter as quadratic costs we have

$$\begin{aligned}H(q) = &\psi^N q + \big(\alpha q + (1-\alpha)(1-q)\big) f(1-r)\\&+ r\big(\alpha q^2 + (1-\alpha)(1-q)^2\big) - r^2 q(1-q)\big(\beta(1-q) + (1-\beta)q\big).\end{aligned}$$

The second derivative of this function is given by

$$H''(q) = 2r + 2r^2(\beta(1-q) + (1-\beta)q - (1-2q)(1-2\beta)).$$

Observe for all $\beta, q \in [0,1]$

$$\beta(1-q) + (1-\beta)q - (1-2q)(1-2\beta) \geq -1.$$

Hence,

$$H''(q) \geq 2r - 2r^2 = 2r(1-r) \geq 0.$$

Therefore, $H(q)$ does not have a maximum in $[0,1]$.

E.3 Comparative statics in the quadratic case

Proposition 22 *When protests are quadratic, comparative statics for threshold $\bar{q}_0$ are as follows*

- *It holds*
$$\frac{\partial}{\partial \psi^N}\bar{q}_0 < 0, \qquad \frac{\partial}{\partial \alpha}\bar{q}_0 < 0, \qquad \frac{\partial}{\partial \beta}\bar{q}_0 < 0.$$

- $\alpha \geq \frac{1}{2}$ *if and only if*
$$\frac{\partial}{\partial \theta}\bar{q}_0 \leq 0, \qquad \frac{\partial}{\partial \sigma}\bar{q}_0 \leq 0.$$

- *If* $\psi^N + (2\alpha - 1) f(1-r) - r^2(1-\alpha+\beta) > 0$
$$\frac{\partial}{\partial \rho}\bar{q}_0 >$$

- *If* $\psi^N + (2\alpha - 1) f(1-r) + r^2(\alpha+1-\beta) < 0$*, then*
$$\frac{\partial}{\partial \rho}\bar{q}_0 < 0$$

Proof. It follows the same argument as in the comparative statics for the linear case. See Supplementary Appendix D. ■

F Technical details for Section 6

F.1 Microfoundations of political parties' objective function

We follow the probabilistic voting model with majority voting and aggregate uncertainty proposed by Persson and Tabellini (2000) based on Lindbeck and Weibull (1987). Recall that parties A and B make simultaneous announcements δ^A and δ^B in every period, with full commitment. Voters are myopic, in the sense that they only value policies according to their utility in period t.[7] Voter j in group i votes for A if

$$U^i(\delta^A) > U^i(\delta^B) + \sigma^{ij} + \mu,$$

where σ^{ij} measures ideological idiosyncratic preference toward party B. σ^{ij} is i.i.d. and drawn from a uniform distribution $\mathcal{U}\left[\frac{-1}{2\phi^i}, \frac{1}{2\phi^i}\right]$. Note the distributions have density ϕ^i and neither group is biased on average toward one of the parties. We could think about this parameter as reflecting another policy dimension orthogonal to policy δ_t, for which political parties cannot make credible commitments but on which

[7]Concretely, voters do not internalize the effect of their choices on the dynamics of identities.

they implement some policy after the election in accordance with their ideology. In a sense, it is a measure of ideological homogeneity within the group that translates into political strength. μ captures average relative popularity of party B, drawn i.i.d. from $\mathscr{U}\left[\frac{-1}{2}, \frac{1}{2}\right]$. Note that without introducing aggregate uncertainty (given by the value of μ), the probability of winning that we define below is not continuous on the announcement, and the model collapses to a modified version of the Downsian model in which all that matters are the preferences of the swing voter. In that case, any forward-looking motive will have no bite, as any party deviating from the preferences of the swing voter losses the elections with probability 1. In that case, the only possible equilibrium is to play the optimal strategy of the static game.[8]

The probability that a randomly drawn voter of group i votes for A is given by

$$Pr(\sigma^{ij} < U^i(\delta^A) - U^i(\delta^B) - \mu) = F^i\Big(U^i(\delta^A) - U^i(\delta^B) - \mu\Big) = \frac{1}{2} + \phi^i[U^i(\delta^A) - U^i(\delta^B) - \mu].$$

Hence, the vote share for party A for policy announcements δ^A and δ^B for given q at time t is

$$\pi^A(\delta^A, \delta^B, q) = \frac{1}{2} + q\phi^N\Big[U^N(\delta^A) - U^N(\delta^B) - \mu\Big] + (1-q)\phi^R\Big[U^R(\delta^A) - U^R(\delta^B) - \mu\Big].$$

We assume a majority voting electoral rule, so party A wins the election at time t if $\pi^A > \frac{1}{2}$. Because at the time announcements are made the popularity shock μ is unknown, π^A is a random variable and therefore party A wins the election with probability p^A given by

$$\begin{aligned}
p^A(\delta^A, \delta^B, q_t) &= Pr\left(\pi^A > \frac{1}{2}\right) \\
&= \frac{1}{2} + \frac{q_t\phi^N\big[U^N(\delta^A) - U^N(\delta^B)\big] + (1-q_t)\phi^R\big[U^R(\delta^A) - U^R(\delta^B)\big]}{q_t\phi^N + (1-q_t)\phi^R} \\
&= \frac{1}{2} + \Phi(q_t)(\delta^A - \delta^B),
\end{aligned}$$

where $\Phi(q) = \frac{(1-q_t)\phi^R - q_t\phi^N}{q_t\phi^N + (1-q_t)\phi^R}$. It follows that party B wins the elections with probability $p^B(\delta^A, \delta^B, q_t) = 1 - p^A(\delta^A, \delta^B, q_t)$.

[8] The results of this section remain if instead of introducing aggregate uncertainty and majority voting we assume that there is no aggregate uncertainty but: a) the benefits from office for each party are proportional to its vote share and; b) the policy implemented is a weighted average of the announcements. This specification yields an equivalent game and it allows to discuss how the degree of proportionality of the electoral system (i.e. how vote shares translate into power shares) affects nation-building prospects.

F.2 Static electoral competition

First, we consider parties that are myopic, in the sense that they do not internalize identity dynamics.[9] Therefore, in each period they solve the static political-economy game, i.e., they maximize the objective function taking what the other party does as given. The Nash equilibria of the static electoral-competition game are characterized by

$$\delta^{*i}(q) = \arg\max_{\delta\in[0,1]} \psi^N q + p^i(\delta, \delta^{*-i}) = \arg\max_{\delta\in[0,1]} \psi^N q + \frac{1}{2} + \Phi(q)(\delta - \delta^{*-i}),$$
$$\delta^{*-i}(q) = \arg\max_{\delta\in[0,1]} \psi^N q + p^{-i}(\delta^{*i}, \delta) = \arg\max_{\delta\in[0,1]} \psi^N q + \frac{1}{2} + \Phi(q)(\delta - \delta^{*i}).$$

It is easy to see that for given q, the symmetric Nash equilibrium is characterized by

$$\delta^{i*}(q) = \arg\max_{\delta} \Phi(q)(\delta - \delta^{-*i}) = \begin{cases} 1 & \text{if } \Phi(q) > 0 \\ [0,1] & \text{if } \Phi(q) = 0 \\ 0 & \text{if } \Phi(q) < 0. \end{cases}$$

Because $\Phi(q)$ is strictly decreasing in q, the previous equilibrium strategy can be equivalently defined as

$$\delta^{i*}(q) = \begin{cases} 1 & \text{if } q_0 < \tilde{q}_S \\ [0,1] & \text{if } q_0 = \tilde{q}_S \\ 0 & \text{if } q_0 > \tilde{q}_S, \end{cases}$$

where $\tilde{q}_S$ is given by $\Phi(\tilde{q}_S) = 0$, that is $\tilde{q}_S = \frac{\phi^R}{\phi^R + \phi^N} \in [0,1]$. Given these equilibrium policies, if $q_0 < \tilde{q}_S$, q_t decreases over time converging to $q = 0$. Alternatively, if $q_0 > \tilde{q}_S$, q_t increases over time converging to $q = 1$. When a group of voters is more concerned about policy δ, in the sense that they are more responsive to changes in the announcement (i.e. higher value of ϕ^i), they are more likely to win elections and, eventually, become the only group in society. Therefore, as in the dynamic game, the survival of regional identities is more likely when the regionalist are demographically big, when the peripheral region is sufficiently pivotal, and when citizens in the regionalist group are ideologically motivated toward identity policy δ with respect to other policy dimensions.

F.3 Parties with opposite nation-building motives

In the electoral competition game we have assumed that both parties want to promote the same national identity. However, as the recent histories of some countries in Africa and Asia show, there are several cases that are better modelled as a game between two *forward-looking* parties that are biased in opposite directions. Unfortunately, charac-

[9] Equivalently, we can have political parties that only live for one period.

terizing the solution to this differential game is technically intractable with the tools developed in this paper, because we cannot restrict our attention to symmetric equilibria. Solving it is a very interesting venue for future research, and it may potentially generate persistent conflict and diversity *as an equilibrium outcome*. However, we believe that also in this case it would be unlikely to obtain either cycles or heterogeneous steady-states, because the two key ingredients for long-run homogeneity under electoral competition (a strong conflict over scarce resources as well as a policy implementation that favours the majority) remain valid in the case where parties have opposite nation-building motives.

In order to sketch how the results could change with parties that represent only the interests of their own groups, we can analyze an example in which parties are shortsighted. However, note that in the shortsighted case the nation-building motive plays no role, so whether parties are biased towards increasing the size of the group with the national or the regional identity is irrelevant. Therefore, in order to have some action, we need to consider political parties that are ideologically motivated to implement some policy. For this, consider a simple modification to the current model, where party A chooses $\delta_t^A = 0$ whenever it wins elections and party B chooses $\delta_t^A = 1$.[10] As in the benchmark model, we assume that there is an idiosyncratic shock and a common shock to party popularity, but the latter is now distributed as a uniform $\mathscr{U}[-1,1]$.[11] Therefore, following the steps above, the probability that party A wins the election at time t, when parties announce $\delta_t^A = 0$ and $\delta_t^B = 1$, is given by

$$p_t^A = \frac{q\phi^N}{q\phi^N + (1-q)\phi^R}$$

Hence, for ideologically motivated parties the probability of winning elections is increasing in the size of the group that it favors with its policies. Recall that dynamics are given by

$$\dot{q}_t = g(q_t) = \begin{cases} q_t(1-q_t)^2 & \text{with prob. } p_t^A \\ -q_t^2(1-q_t) & \text{with prob. } 1-p_t^A \end{cases}$$

As compared to our model of electoral competition, the policy announcements of candidates do not converge because of their extreme ideological bias. As a result, q does not always move in the same direction and the system does not necessarily reach a homogeneous steady state. However, if enough time passes, we should expect q eventually moving in the same direction. The reason is that the biggest group has a higher probability of winning elections and, as a result, get its desired policy. This increases the size of this group through the cultural evolution mechanism, which in turn makes

[10] This example corresponds to a situation where parties are ideologically motivated and cannot commit to implement other policies once they are in office. Despite its simplicity, this assumption captures well the situation of countries such as Nigeria and Kenya, where parties are generally shortsighted, represent different ethnic groups and take turns in power to loot the country.

[11] This change is just to make probabilities bounded between 0 and 1.

them more likely to win elections again. Therefore, even with shortsighted and ideologically motivated candidates, homogeneous populations are the most likely long run outcome, because majority groups tend to become larger over time.

F.4 Voters in the central region

In this subsection we show that introducing voters in the central region does not qualitatively change the results of the electoral-competition game. The reason is that including these voters only changes the function $\Phi(q)$. Therefore, the key properties of the objective function of the central government remain similar and the key features needed for the proof go through. The main qualitative difference comes from the fact that, for some regions of parameters, some trivial cases might arise in which $\Phi(q)$ is lower than zero for all q. We illustrate this last point by means of an example.

Assume the central government is democratically elected each period by people of the central and peripheral regions. The country as a whole has a population of size 1, out of which a fraction $\lambda \in (0,1)$ lives in the peripheral region and a fraction $1-\lambda$ lives in the central region. Within the peripheral region, a fraction q belong to group N and a fraction $1-q$ belong to group R. Utilities are given by[12]

$$\begin{aligned} U^N(\delta) &= g^N = 1-\delta \\ U^R(\delta) &= g^R = \delta \\ U^C(\delta) &= g^N = 1-\delta. \end{aligned}$$

Here, we have made the simple but natural assumption that voters in the central region have the same preferences as nationalist individuals in the peripheral region. This aims to capture the idea that citizens in the central region are socialized to the national identity and enjoy the nationalist public good in the same way as nationalist individuals in the peripheral region. As before, voter j in group i votes for A if

$$U^i(\delta^A) > U^i(\delta^B) + \sigma^{ij} + \mu.$$

Assuming majority voting as before

$$p^A(\delta^A, \delta^B, q) = \frac{1}{2} + \Phi(q)(\delta^A - \delta^B),$$

where $\Phi(q)$ is now given by

$$\Phi(q) = \frac{\phi^R \lambda(1-q) - \phi^N \lambda q - \phi^C(1-\lambda)}{\phi^N \lambda q + \phi^R \lambda(1-q) + \phi^C(1-\lambda)}.$$

When $\phi^R\lambda - \phi^C(1-\lambda) < 0$, we have $\Phi(q) < 0, \forall q \in [0,1]$. This is the main difference with respect to the previous case, where for any value of the parameters it was guaranteed that $\Phi(q)$ always took positive and negative value in q between 0 and 1. Now,

[12] To simplify on parameters, we assume that the total budget of the government is of size 1.

if $\phi^R\lambda - \phi^C(1-\lambda) < 0$, the only equilibrium is given by $\delta = 0$ for any initial q_0. This corresponds to the case in which the regionalist group is not sufficiently pivotal in national elections, which happens when the region is sufficiently small (low λ) or when citizens in the central region are relatively more ideologically concerned with policy δ than the regionalists (ϕ^C relatively large with respect to ϕ^R). In the non-trivial case when $\phi^R\lambda - \phi^C(1-\lambda) > 0$, we are back to a similar case where voters in the central region are not introduced. Therefore, the equilibrium of the dynamic game is analogous, with the minor difference that the thresholds $\tilde{q}_S$ and $\tilde{q}_D$ are additionally affected by the parameters λ and ϕ^C.

One could think of more realistic and detailed specifications that would yield more interesting comparative statics with respect to the two thresholds, without changing the method of the proof for the results in Section 6. For instance, we could have specified that citizens in both regions experiment disutility from protests. This could lead to a case where some citizens in the central region might vote for a policy that favors regionalist individuals, because their desire to reduce conflict might offset their nationalist sentiment. In this case, the persistence of regional identities in democracies would be a function of the complex interaction between the ideological concerns about identity policies of the three groups (ϕ^i), the size/pivotality of the peripheral region (λ) and parameters capturing the impact of protests.

F.5 Proof of Proposition 7

Proposition 7 *The threshold $\tilde{q}_D$ is decreasing in ψ^N*

$$\frac{\partial}{\partial \psi^N}\tilde{q}_D \leq 0,$$

with limiting cases

$$\lim_{\psi^N \to 0} \tilde{q}_D = \tilde{q}_S, \qquad \lim_{\psi^N \to \infty} \tilde{q}_D = 0.$$

On the contrary, $\tilde{q}_D$ is increasing in ρ:

$$\frac{\partial}{\partial \rho}\tilde{q}_D \geq 0.$$

Proof. We first show the comparative statics on the parameter ψ^N. Simply recall the derivative of the recovered value function when $\delta^*(q) = 0$

$$V_q(q) = \psi^N\left[\frac{2}{\rho} + \frac{1}{q(1-q)^2 m(q)}\int_q^1 m(x)dx\right]$$

Observe that $V_q(q)$ is strictly increasing in ψ^N, hence if we have $\psi^N < \psi^{N'}$ and for some $q < \tilde{q}_S$ we have that

$$V_q(q;\psi^N) > 2\frac{\Phi(q)}{q(1-q)},$$

then it also follows that

$$V_q(q;\psi^{N'}) > V_q(q;\psi^N) > 2\frac{\Phi(q)}{q(1-q)}.$$

This means that if $q < \tilde{q}_D(\psi^N)$, then $q < \tilde{q}_D(\psi^{N'})$ too, and hence

$$\tilde{q}_D(\psi^{N'}) \le \tilde{q}_D(\psi^N)$$

Moreover, it is easy to see that

$$\lim_{\psi^N\to 0} \tilde{q}_D(\psi^N) = \tilde{q}_S, \qquad \lim_{\psi^N\to\infty} \tilde{q}_D(\psi^N) = 0.$$

For the discount factor ρ, the comparative statics are proven using a similar argument.
■

G Endogenous tax rate

We modify the baseline model such that the government is able to choose the tax rate $\{r_t\}_{t\ge 0}$ as well as the relative provision of each type of public good $\{\delta_t\}_{t\ge 0}$. The resulting government's problem has two control variables and is given by

$$\begin{aligned} \max_{r_t,\delta_t,\in[0,1],\forall t\ge 0} \int_0^\infty & e^{-\rho t} W(q_t,\delta_t,r_t)\mathrm{d}t \\ \text{s.t.} \quad & \dot{q}_t = r_t q_t(1-q_t)(1-\delta_t-q_t) \\ & q(0)=q_0,\ q_t\in[0,1]. \end{aligned} \tag{50}$$

with corresponding HJB equation given by

$$\rho V(q) = \max_{r,\delta} W(q,\delta,r) + g(q,r,\delta)V'(q), \tag{51}$$

where

$$\begin{aligned} W(q,\delta,r) =& \psi^N q_t + \alpha q\big(f(1-r)+(1-\delta)r\big) + (1-\alpha)(1-q)\big(f(1-r)+\delta r\big) \\ & - r\big(\beta q\delta + (1-\beta)(1-q)(1-\delta)\big) \\ g(q,\delta,r) =& rq(1-q)(1-q-\delta). \end{aligned}$$

The following proposition holds:

Proposition 23 *Assume utility from private consumption is $f(x)=\theta\frac{x^{1-\sigma}}{1-\sigma}$ with $\theta,\sigma\in(0,1)$. Then, open neighborhoods of $q=0$ and $q=1$ in $[0,1]$ exist, say, $\mathcal{O}(0)$ and $\mathcal{O}(1)$, such that*

$$r^*(q)>0,\ \textit{with}\ \delta^*(q)=1\ \forall q\in\mathcal{O}(0), \quad \textit{and} \quad r^*(q)>0,\ \textit{with}\ \delta^*(q)=0\ \forall q\in\mathcal{O}(1).$$

Proof. From 51, the optimal tax-rate for $q=0$ and $q=1$ is given by

$$r^*(0)=r^*(1)=1-\theta^{\frac{1}{\sigma}}\in(0,1),$$

with corresponding value function

$$\rho V(0) = (1-\alpha)\big(1+\theta^{\frac{1}{\sigma}}\frac{\sigma}{1-\sigma}\big) \quad \rho V(1) = \psi^N + \alpha\big(1+\theta^{\frac{1}{\sigma}}\frac{\sigma}{1-\sigma}\big).$$

First, we prove that $r^*(q)$ is continuous at $q=1$ and at $q=0$, by contradiction. Assume not, so $\lim_{q\to 1} r^*(q) = c \neq 1-\theta^{\frac{1}{\sigma}}$. From Theorem 1 we know $\delta^*(q)$ is continuous at $q=0$ and at $q=1$. Then, by continuity of $V(q)$ it must hold

$$\lim_{q\to 1}\rho V(q) = \psi^N + \alpha(f(1-c)+c) = \psi^N + \alpha\big(1+\theta^{\frac{1}{\sigma}}\frac{\sigma}{1-\sigma}\big) = \rho V(1),$$

which implies

$$f(1-c)+c = f(\theta^{\frac{1}{\sigma}}) + 1 - \theta^{\frac{1}{\sigma}} = \max_x f(1-x)+x.$$

Observe the function $f(1-x)+x$ is strictly concave, so the only solution of the previous equation is precisely $c = 1-\theta^{\frac{1}{\sigma}}$, and therefore

$$\lim_{q\to 1} r^*(q) = 1-\theta^{\frac{1}{\sigma}} = r^*(1),$$

which proves $r^*(q)$ is continuous at $q=1$. Similarly for $q=0$. Because $r^*(1) = r^*(0) > 0$, by continuity of $r^*(q)$ open neighborhoods in $[0,1]$ of $q=0$ and $q=1$ exist such that $r^*(q) > 0$ for all q in those neighborhoods.

For the second part of the proposition, we use continuity of $\delta^*(q)$ at $q=0$ and at $q=1$, which follows from Theorem 1. By continuity of $r^*(q)$ at $q=0$ and $q=1$, we can find open neighborhoods of $q=1$ and $q=0$, $\mathcal{O}(0)$ and $\mathcal{O}(1)$ respectively, such that $r^*(q)$ and $\delta^*(q)$ are continuous inside them. Also, from Theorem 1, either $\delta^*(q)=0$ or $\delta^*(q)=1$, with $\delta^*(0)=1$ and $\delta^*(1)=0$. By continuity of $\delta^*(q)$ in $\mathcal{O}(0)$ and $\mathcal{O}(1)$, it follows

$$r^*(q) > 0,\ \delta^*(q) = 1 \quad \forall q \in \mathcal{O}(0), \quad \text{and} \quad r^*(q) > 0,\ \delta^*(q) = 0 \quad \forall q \in \mathcal{O}(1).$$

■

The previous result implies that when the population is largely homogeneous, it is better for the government to collect taxes, provide public goods, and homogenize toward the prevailing identity, because at those states the participation rate in protests of the minority group is small and it is optimal to pursue full homogenization.

G.0.1 Toward a general solution

Unfortunately, finding a closed-form solution of the optimal tax rate r is analytically intractable given the cubic law of motion of the state variable q. However, in this section we outline the steps toward a full solution of problem 50.

First, we show that the solution to problem 50 is equivalent to a sequential maximization problem. From Theorem 1, we know that the solution $\delta^*(r,q)$ for any r and

q is given by

$$\delta^*(r,q) = \arg\max_{\delta} W(q,\delta,r) + g(q,r,\delta)V'(q) = \begin{cases} 1 & \text{if } q \leq \bar{q}_0(r) \\ 0 & \text{if } q \geq \bar{q}_0(r). \end{cases}$$

for any given r, and q. That is, for any r, including the optimal tax-rate $r^*(q)$, we know that $\delta^*(r^*(q), q)$ can only take two values, i.e. $\delta^*(r^*(q), q) \in \{0,1\}$ for all $q \in [0,1]$. The previous result greatly simplifies problem 50, to

$$\rho V(q) = \max_{\delta \in \{0,1\}} \left\{ \max_{r \in [0,1]} W(q,0,r) + g(q,0,r)V'(q), \max_{r \in [0,1]} W(q,1,r) + g(q,1,r)V'(q) \right\} \quad (52)$$

Next, to find interior solutions $r^*(q) \in (0,1)$ we could solve each sub-problem in problem 52 by solving the corresponding ODE obtained from the envelope and first order conditions of the HJB equation. However, there are no analytic solutions to those ODEs. To illustrate this point, we can look at the solution for low values of q, for which we know $\delta^*(q) = 1$, and hence the corresponding ODE for $r^*(q)$ is given by

$$r_q^1 = \frac{1}{r f''(1-r)(\alpha q + (1-\alpha)(1-q))} \Big\{ \psi^N + (2\alpha - 1)\big(f'(1-r)r + f(1-r)\big) \\ \frac{\rho}{q^2(1-q)}\Big((\alpha q + (1-\alpha)(1-q))f'(1-r) \\ -(1-\alpha)(1-q) + \beta q\Big)\Big\}.$$

However, the previous ODE does not have a closed-form analytic solution, even after choosing specific values of the parameters α, β, and ρ and σ. Therefore, obtaining a full complete characterization of $r^*(q)$ is analytically intractable. Similarly, the corresponding ODE for large values of q is given by

$$r_q^0 = \frac{1}{r f''(1-r)(\alpha q + (1-\alpha)(1-q))} \Big\{ \psi^N + (2\alpha - 1)\big(f'(1-r)r + f(1-r)\big) \\ + \frac{\rho}{q(1-q)^2}\Big(-(\alpha q + (1-\alpha)(1-q))f'(1-r) \\ + (\alpha q - (1-\beta)(1-q))\Big)\Big\}$$

Furthermore, observe how general results about the monotonicity of $r^*(q)$ are difficult to obtain because the sign of the previous ODEs depend on the other parameters of the model. Hence different parameter combinations will lead to different results.

For illustrative purposes, we numerically solve problem 52 using the numerical methods proposed in Achdou et al. (2017). Figures 16 and 17 illustrate the solution for $r^*(q)$ and $\delta^*(q)$ for some parameters. We can see that whenever $r^*(q) > 0$ for all $q \in [0,1]$, the optimal solution for $\delta^*(q)$ resembles the bang-bang nature of the baseline model; that is, a threshold $\bar{q}_0$ exists such that $\delta_t = 1 \ \forall t$, with q_t converging to

Parameter values: $\psi^N = 0.5$, $\alpha = 0.5, \beta = 0.5, \theta = 0.2, \sigma = 0.5$, $\rho = 0.5$

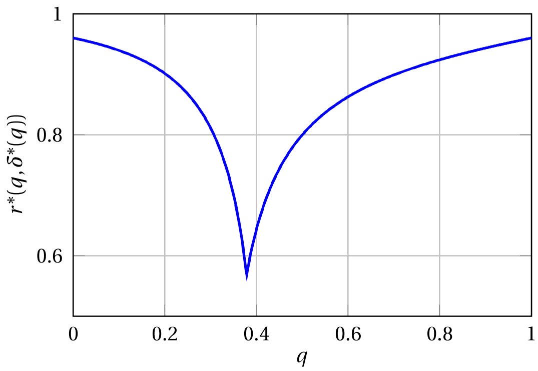

Figure 16: $r^*(q)$

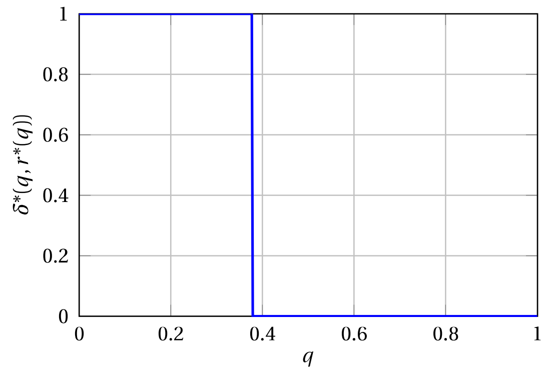

Figure 17: $\delta^*(q)$

$q = 0$ whenever $q \leq \bar{q}_0$ and vice versa. All numerical examples display similar qualitative results. Importantly, even when r is chosen optimally (and conditional on being strictly positive), convergence to a extreme steady state still occurs, showing softer budget constraints do not eliminate the overall conflict.[13]

Moreover, the numerical solution suggests the optimal tax rate $r^*(q)$ is higher for more homogeneous populations and reaches a minimum at the indifference threshold $\bar{q}_0$, as a result of a *static* trade-off present in the choice of r. On the one hand, an increase in r reduces the private consumption of both groups. On the other hand, it increases the resources available to provide one of the two public goods. For intermediate values of q, the negative effect dominates because all citizens are affected by the tax collection, but only one group benefits from public-good provision. However, as the government comes closer to the homogeneous states, the positive effects dominate because the benefits from the public-good provision are larger. Moreover, we can see r increases sharply at early stages and at diminishing rate afterwards. This behavior results from the *dynamic* effect of changing r and directly affects the law of motion: By increasing r, the government can move faster in any direction. Therefore, for intermediate values of q, the government wants to change r sharply in order to rapidly reduce the size of the group that pushes welfare down.

[13] For some parameter choices, we found cases in which $r^*(q) = 0$ for intermediate values of q. This finding corresponds to cases in which the government is sufficiently welfarist and individuals' marginal utility of private consumption is relatively big (sufficiently high σ and θ).