The Tocqueville Paradox:
When Does Reform Provoke Rebellion?*

Evgeny Finkel               Scott Gehlbach
Johns Hopkins SAIS          University of Wisconsin–Madison

This version: November 4, 2018

Abstract
We develop a model of reform and rebellion to explore Alexis de Tocqueville’s conjecture
that reform provokes political unrest. Our theory emphasizes that reform often must be
implemented by local actors with a stake in the status quo. In this setting, the promise
of reform represents an implicit contract against which subsequent implementation
is measured: when implementation falls short of the promise, citizens are aggrieved
and more likely to rebel. In equilibrium, when reform is predominantly under local
control, a more ambitious reform encourages greater implementation; nonetheless, the
equilibrium probability of rebellion also increases. We illustrate our argument with a
discussion of Russia’s Emancipation Reform of 1861.

*For helpful comments, we thank Avi Acharya, Kate Baldwin, Chris Blattman, John Earle, Bob Gibbons,
Florian Hollenbach, Mike Miller, Steve Nafziger, Jack Paine, Chris Price, Emily Sellars, Mehdi Shadmehr,
Konstantin Sonin, and David Weimer; participants in the annual meetings of APSA, ASEEES, SIOE, SPSA;
and participants in seminars at Chicago, NYU, Texas A&M, Virginia, Wisconsin, and Yale. Gabriel Martinez
and Julian Waller provided excellent research assistance.
The regime that a revolution destroys is almost always better than the one that immediately preceded it, and experience teaches that the most dangerous time for a bad government is usually when it begins to reform.

---

Alexis de Tocqueville ([1856] 2011, p. 157)

Contemporary models of political economy posit an intuitive relationship between reform and rebellion. By transferring utility to some excluded group, reform raises the opportunity cost of rebellion, thus reducing the probability of unrest. This logic drives numerous models of regime change and autocratic stability (e.g., Boix, 2003; Acemoglu and Robinson, 2006; Gandhi and Przeworski, 2006; Dunning, 2008; Ansell and Samuels, 2010; Svolik, 2012; Miller, 2013), reflecting the reality in authoritarian states that rebellion is often the only recourse to bad government.

This view of the relationship between reform and rebellion is entirely at odds with that famously articulated by Alexis de Tocqueville. As captured in the epigraph to this paper, Tocqueville argued that reforming governments were often at greatest risk of revolution. Tocqueville’s argument is multifaceted, but at its core it emphasizes the role of reform in raising expectations that cannot be satisfied. This perspective finds echoes in important work on reform and rebellion that predates the models cited above. Key contributions include Davies’ (1962) famous “J-curve” theory of revolution, in which a period of wealth and advancement is followed by backsliding and worsening of conditions, eventually leading to revolution; Gurr’s (1970) theory of relative deprivation, in which expectations of improvement are juxtaposed with a much harsher reality, leading “men” to rebel; and Huntington’s (1968) observation that reform may be a “catalyst” rather than “substitute” for social instability.

Tocqueville’s argument that unfulfilled expectations produce unrest is intuitive, but why

---

1A related phenomenon is the emergence of leftist insurgencies in left-leaning democracies; see Staniland (2018).
cannot the expectations raised by reform be fulfilled? One possibility is that reform, once announced—and expectations raised—can still be blocked by actors with a stake in the status quo. Expressing this view, Oberschall (1995, pp. 155–157) suggests that “Tocqueville’s paradox” is in fact hardly surprising, as “reform attempts tend to be half-hearted and ineffective because of institutional weakness and deliberate efforts by certain groups to undermine them.” The actual outcome is “erratic reformism,” exacerbating rather than quieting discontent.

This perspective, in turn, raises further questions. To what extent do the “certain groups” in a position to block implementation of reform internalize the resulting rebellion? Is there always a tradeoff between stability and reform, or is it possible to pursue the latter without risking the former? Finally, is reform indeed most dangerous for what Tocqueville referred to as “bad governments”?

To answer these questions, we develop a simple model of reform and rebellion. Our modeling approach incorporates two key assumptions. First, we assume that some part of reform, modeled simply as the reallocation of a divisible resource, must be implemented by a strategic agent—a local elite, in our formalization—with a stake in the status quo. Even autocratic rulers cannot rule by fiat alone, but instead must rely upon formal or informal agents to carry out bureaucratically demanding tasks. Those agents with sufficient expertise to implement reform—or sufficient power to block its local implementation—are often those with the most to lose.

An instructive case, which we discuss at length below, is Russia’s emancipation of the serfs in 1861: implementation of this complicated reform was by necessity substantially delegated to the local nobility that possessed the land on which serfs resided. More generally, land reform often provides opportunities for obstruction by local elites (Albertus, 2015; Albertus and Kaplan, 2012). When in the wake of decolonization numerous African states embarked on heavily “top-down” land reforms, the results were often abysmal. Recognizing these failures, in the early 1990s central states “backed off” and delegated the implementation of
land redistribution and titling to local actors (Bruce and Knox, 2009). Such processes also extend to other types of reform. When the Ottoman Empire moved from tax farming to direct taxation as part of the mid-19th century Tanzimat reforms, it often had to rely on pre-reform tax farmers to implement the change, even though these individuals had everything to lose and almost nothing to gain from the new situation (Aytekin, 2012). Similarly, the Industrial Communities program in Juan Velasco’s Peru, which sought to increase worker control and profit sharing in industrial enterprises, could only succeed with the acquiescence of enterprise managers (Stokes, 1995, pp. 33-4). Even the most capacious states often depend on local agents to implement reforms. For instance, education reforms typically depend on the conduct of teachers, who are “the final brokers when it comes to implementing policy” in their classrooms (Spillane, 2000, p. 142), regardless of whether they gain or lose from reform.

Our second key assumption is that reform represents an implicit contract against which subsequent implementation is measured. To the extent that implementation falls short of the promise of reform, citizens are aggrieved and consequently more likely to rebel. This perspective builds most directly on recent work in economics on contracts as reference points—with the key difference that the contract is imposed on, rather than negotiated by, the actor responsible for its implementation. As in that literature (see especially Hart and Moore, 2008; Hart and Holmstrom, 2010), we adopt the stark assumption that citizens’ aggrievement, and thus propensity to rebel, is increasing in the difference between the maximum they could have received (the promise of reform) and what they actually do (the implementation of reform).

The verisimilitude of this behavioral assumption is plausibly supported by related work on two distinct phenomena. First, a long literature in various disciplines emphasizes that rebellion is driven, at least in part, by “grievances” or “expressive” concerns (e.g., Gurr, 1970; Scott, 1976; Wood, 2003), sometimes in interaction with more instrumental motivations (van Zomeren et al., 2004; van Zomeren, Leach and Spears, 2012; Humphreys and Weinstein,
In our setting, such grievances are driven by disappointment with the implementation of reform. Second, a substantial literature demonstrates the importance across a range of field and laboratory environments of “reference dependence”—that is, the tendency of individuals to “normally perceive outcomes as gains and losses...relative to some neutral reference point” (Kahneman and Tversky 1979, 274). We argue that the promise of reform can serve as one such reference point.

As our discussion of the Russian case below illustrates, the promise of reform is often quite visible, creating the potential for disappointment—and thus unrest—when the implementation of reform is blocked. Intuitively, we can think of citizens as “endowed” by reform with a share of the contested resource (Thaler, 1980; Knetsch, 1989; Kahneman, Knetsch and Thaler, 1990); the failure of elites to fulfill that promise is experienced as a loss of the endowment. At the same time, such grievances may be at least partially offset to the extent

2Chong (2014, ch. 4) refers to expressive behavior as noninstrumental but “narrowly rational,” which accords with our assumption that such concerns enter the cost-benefit calculation in the decision to rebel. Blattman and Miguel (2010) suggest that the weak statistical evidence for such motivations in some studies may be a consequence of the crude proxies employed, and they call for the development of better measures of grievances. For a formal theory of grievance-based (vs. political-process) contentious politics, see Shadmehr (2014).

3For reviews, see Camerer, Loewenstein and Rabin (2004) and DellaVigna (2009). Formal models of political behavior have considered a variety of reference points, including the status quo (Alesina and Passarelli, 2015; Lockwood and Rockey, 2015), rational expectations of future payoffs (Grillo, 2016; Acharya and Grillo, 2017), and aspirations established by parents (Besley, 2016); see also below.

4The assumption that citizens are more likely to rebel when implementation falls short of the promise of reform can also be motivated by the observation that individuals are willing to take costly actions to punish those who hurt them (Rabin, 1993) and by evidence that collective action is more likely when citizens are able to trace blame to specific, identifiable actors.
that reform improves on the status quo—another important reference point that is empha-
sized in some experimental work.\(^5\) Rather than choosing between these two reference points
in an ad hoc manner, we parameterize their relative importance. Our key results relate to
the weight that citizens place on gains over the status quo (versus losses from incomplete
implementation), relative to the extent to which the implementation of reform is under local
(rather than central) control.

Our behavioral assumptions pin down the probability of unrest, \textit{given} the degree of local
implementation. The need for a model is to evaluate the strategic response of local elites, for
whom we assume rebellion is costly. We proceed in steps. We first present a simple model
with a single elite and unitary citizenry. For the sake of transparency, we ignore instrumental
motivations to rebel. This model produces a sharp prediction: When the implementation
of reform is predominantly under local control, then a more ambitious reform always leads
to greater implementation by local elites; at the same time, the equilibrium probability of
rebellion also increases. Unrest is the price of cooperation by local elites.

We subsequently show that these results are robust to relaxing various assumptions of
our simple model. First, we demonstrate that the assumption of a unitary citizenry can
be rationalized with a simple model of collective action among citizens of that locality.
Second, we assume that the citizenry may be motivated by instrumental as well as expressive
concerns. Third, we extend the model to allow for bargaining by multiple local elites and
citizenries, where the expressive benefit to any citizenry from rebelling is increasing in the
number of other localities that rebel. Across all these generalizations and extensions, our

\(^5\)For discussion, see K˝ oszegi and Rabin (2006). Our formulation of dual reference points
is consistent with the “reference lotteries” in K˝ oszegi and Rabin if one interprets the status
quo and promise of reform as potential outcomes, though our analysis does not invoke that
paper’s “extreme” (p. 1135) assumption that expectations under this lottery are determined
by equilibrium strategies.
results are robust: if control over the implementation of reform is predominantly local, then—notwithstanding the strategic response of local elites—reform necessarily increases the equilibrium probability of rebellion.

Our analysis concludes with a discussion of various positive and normative implications of our model. We examine the conditions under which reform is most likely to provoke rebellion: when state capacity is low, or when the nature of reform necessitates local implementation. Under these conditions, reformist governments face a tradeoff: a more ambitious reform implies greater aggregate (central and local) implementation, but also a greater risk of rebellion. We illustrate these and other insights of the model with a discussion of an important case of reform and rebellion: the emancipation of the serfs in Russia in 1861.

Our theoretical approach leans heavily on reference dependence, one of the pillars of the “prospect theory” of Kahneman and Tversky (1979). Although increasingly common in neighboring disciplines, applications of prospect theory are relatively unusual in political science outside the field of international relations, and they are rare in the study of protest and political violence (but see Fanis, 2004; Tezcür, 2016). One recent paper in the economics literature, however, does formally explore collective action in the context of reference dependence: Passarelli and Tabellini (2017), which models the relationship between “emotions” (analogous to expressive motivations in our setting) and political unrest. Our work shares some similarities and differences with that paper. As in Passarelli and Tabellini, we model psychological grievances as an important determinant of rebellion, though our focus on the decentralized implementation of reform is novel. Moreover, as in that paper, we establish equilibrium uniqueness in the presence of strategic complementarities (here, when considering coordination both within and across localities) directly through preference heterogeneity, not—as in the extensive literature on “global games”—differential information.

---

6 For a review and application to cost-benefit analysis, see Weimer (2017).

7 Levy (2003), McDermott (2004), and Mercer (2005) provide early reviews; see also the various cites above.
(and thus preference heterogeneity). Our approach differs from Passarelli and Tabellini in the specific reference points that determine psychological gains and losses—here, the status quo of no reform and the promise of reform itself; in Passarelli and Tabellini, the solution to a biased social-welfare maximization problem.

Our work also connects to a burgeoning literature in political economy on the causes and consequences of state capacity (e.g., Besley and Persson, 2010). Within political science, a distinction is sometimes drawn between the capacity of principals and that of agents (e.g., legislatures and bureaucracies, respectively; see the discussion in Gehlbach, 2013, ch. 5). Our approach is closest to the former tradition: we ask how the relationship between reform and rebellion depends on the capacity of an (unmodeled) central government to implement reform on its own. Our finding that reform provokes rebellion only when such capacity is weak relates to work by Bertrand (2013) and Haggard and Kaufman (1992, 1995), who argue that strong state institutions reduce the likelihood of pressure from below in the wake of reform.8

Not least, we build on a large literature on the relationship between reform and rebellion, some of which is cited above. A related body of work (Acemoglu, Egorov and Sonin, 2018; Healy, Kosec and Mo, 2017) considers Tocqueville’s varying conjectures about the relationship between social mobility (i.e., the possibility of moving up or down the income distribution) and social stability (de Tocqueville, [1835/40] 2002, [1856] 2011). Our model largely abstracts from such considerations, treating the elite and citizenry within any locality as unitary actors.

8Numerous models consider mechanisms by which autocrats might reduce the risk of rebellion without explicitly tying unrest to reform; see Gehlbach, Sonin and Svolik (2016). For a model of reform and rebellion in which “capacity” takes the form of resolve, see Carroll and Pond (2018).
1 Baseline model and analysis

To build intuition, we analyze a bargaining game featuring an *elite* and a *citizenry*, where the citizenry is motivated only by expressive concerns. It is useful to think of the elite and citizenry as local actors within a larger polity. We ignore for now the role of coordination within and across localities, both of which we take up in the following section.

We are interested in the response of the citizenry to reform $\gamma \in (0, 1]$, where the parameter $\gamma$ denotes the proportion of an infinitely divisible resource to be transferred to, or retained by, the citizenry. We do not provide a model of the process by which $\gamma$ is chosen, though our analysis implies a number of lessons for reform design, which we discuss below. Implicitly, we model the subgame that follows the unmodeled choice of $\gamma$ by a central government.\footnote{In a related context, Gerber, Lupia and McCubbins (2004) analyze the implementation game (involving a legislature and bureaucracy) that follows an exogenous “reform” (citizen initiative).} We let $\hat{\gamma}$ denote the status quo allocation of the resource to the citizenry, where $\hat{\gamma} < \gamma$.

The implementation of reform depends in part on the behavior of the local elite. In particular, we assume exogenous proportion $\lambda \in (0, 1)$ of the resource to be under control of the elite, which chooses a *local implementation* of reform $x \in [0, \gamma]$. In contrast, proportion $(1 - \lambda) \gamma$ is transferred automatically to the citizenry—for example, because an unmodeled central government has sufficient capacity to force that decision on the elite, an intuition to which we return further below. We refer to this latter, automatic transfer as *central implementation*. Many of our key results relate to the degree to which reform is locally implemented, as captured by $\lambda$. Figure 1 provides a graphical illustration of this process.

Following implementation of reform, the citizenry in each locality decides whether to rebel. In the baseline model examined here, the motivation to rebel is based entirely on expressive concerns—not on any anticipated material gains from rebellion. Peering behind the veil of a unitary citizenry, one can rationalize this assumption by assuming that the
material gains from rebellion are non-excludable, whereas the “warm glow” from rebellion is experienced if and only if a citizen participates in rebellion. Further below we examine the robustness of our results to assuming that the citizenry values the contested resource directly.

In particular, we assume that the citizenry is more inclined to rebel, the larger its dissatisfaction with the implementation of reform. Formally, the citizenry compares the payoff from not rebelling, which we normalize to zero, to the expected payoff from rebelling, which we define as

$$\Gamma(x; \gamma) - \epsilon,$$

where $\Gamma(x; \gamma)$ denotes the citizenry’s grievance, which is a function of the endogenous local implementation of reform $x$ and the exogenous reform $\gamma$. The random variable $\epsilon$ represents the (material, psychological, etc.) cost of rebellion, drawn from distribution $F$, that is realized only after the elite chooses $x$. We assume that $F$ is twice differentiable, with density $f$, and strictly increasing on an interval sufficiently wide to ensure that there are realizations
Figure 2: The citizenry’s grievance is increasing in the degree to which local implementation $x$ falls short of the promise of reform $\gamma$ (area with dark shading) and decreasing in the degree to which the implementation of reform—central as well as local—improves on the status quo $\hat{\gamma}$ (area with light shading). The model parameterizes these two evaluations, with weight $1 - \beta$ given to the former and weight $\beta$ given to the latter.

Our key results then further exploit the assumption that the hazard rate,

$$H(s) \equiv \frac{f(s)}{1 - F(s)},$$

is strictly increasing on the support of $F$. This property characterizes a wide class of distributions, including the uniform and normal.

To capture the idea that reform is an incomplete contract against which subsequent implementation is measured, we assume that the citizenry’s grievance is increasing in the degree to which local implementation $x$ falls short of the promise of reform $\gamma$. (By assumption, central implementation is complete.) We also allow for the possibility that reform reduces the citizenry’s grievance to the extent that the implementation of reform—central as well as local—improves on the status quo $\hat{\gamma}$. Figure 2 illustrates these evaluations. We parameterize the relative importance of the two evaluations, such that weight $1 - \beta$ is given to the former and weight $\beta$ given to the latter. Formally, the citizenry’s grievance is
\[
\Gamma (x; \gamma) = \underbrace{(1 - \beta) [\lambda (\gamma - x) + (1 - \lambda) (\gamma - \hat{\gamma})]}_{\text{Underimplementation}} - \underbrace{\beta [\lambda (x - \hat{\gamma}) + (1 - \lambda) (\gamma - \hat{\gamma})]}_{\text{Gains over status quo}}
\]

Simplifying gives

\[
\Gamma (x; \gamma) = \lambda (\gamma - x) - \beta (\gamma - \hat{\gamma}).
\]

The citizenry’s grievance is thus a weighted difference of the degree to which a) local implementation falls short of the promise of reform and b) the promise of reform improves on the status quo, where the weights on the two terms correspond to the degree of local implementation \((\lambda)\) and the relative importance of the status quo in the citizenry’s evaluation of gains and losses from reform \((\beta)\).

Rebellion is costly to the elite. We assume that, in the event of a rebellion, the elite loses proportion \(p \in (0, \gamma)\) of the resource under local control. (We can interpret the parameter \(p\) as the probability that the citizenry gains control of the local resource, which for reasons discussed above does not enter the calculus to rebel; below we demonstrate robustness to relaxing this assumption. Alternatively, \(p\) may represent the destruction of some proportion of the local resource in the event of rebellion.) In contrast, if no rebellion is attempted, the elite transfers \(\lambda x\) to the local citizenry. Finally, as previously assumed, regardless of whether the citizenry rebels, the elite transfers the exogenously mandated share \((1 - \lambda) \gamma\). The elite’s preferences are represented by its expected share of the contested resource,

\[
\Pr (\text{rebellion} \mid x) [\lambda (1 - p) + (1 - \lambda)(1 - \gamma)]
+ [1 - \Pr (\text{rebellion} \mid x)] [\lambda (1 - x) + (1 - \lambda)(1 - \gamma)],
\]

which can be equivalently represented as

\[
-x + \Pr (\text{rebellion} \mid x) (x - p).
\]

Summarizing, the game proceeds as follows: The elite chooses a level of local implementation \(x \in [0, \gamma]\) to maximize Expression 3, following which the random cost of rebellion \(\epsilon\) is
realized. The citizenry then decides whether to rebel, given Expression 1 and Equation 2. The baseline model thus represents a “standard” bargaining game with a take-it-or-leave-it offer and an outside option, but for the fact that the “dissatisfied” party (here, the citizenry) is motivated by expressive rather than instrumental concerns.

Given these assumptions, at an interior solution, the equilibrium level of local implementation is given implicitly by

\[ x^* = p - \frac{1}{\lambda} \cdot \frac{1 - F(\Gamma(x^*; \gamma))}{f(\Gamma(x^*; \gamma))}, \tag{4} \]

which follows directly from the first-order condition for the elite’s problem. (In the online appendix, we show that the second-order condition is satisfied by assumption of an increasing hazard rate.) From this it follows that the equilibrium relationship between local implementation and reform can be expressed as

\[ \frac{\partial x^*}{\partial \gamma} = \alpha (x^*) \cdot \frac{\lambda - \beta}{\lambda}, \tag{5} \]

where

\[ \alpha (x^*) \equiv \left[ 1 + \frac{H'(\Gamma(x^*; \gamma))}{[H(\Gamma(x^*; \gamma))]^2} \right]^{-1} \frac{H'(\Gamma(x^*; \gamma))}{[H(\Gamma(x^*; \gamma))]^2} \]

is a fraction bounded by zero and one that (roughly) captures the relative rate of change in the hazard rate in a neighborhood of the equilibrium grievance. (We suppress from \( \alpha(\cdot) \) the argument \( \gamma \) for notational compactness.) When \( F \) is uniform, \( \alpha (x^*) = \frac{1}{2} \).

This relationship can be interpreted as follows. When reform is predominantly locally implemented (i.e., when \( \lambda > \beta \)), implying that citizens’ expressive motivations—and hence willingness to rebel—are driven by the degree to which local elites fail to deliver on the promise of reform, those elites unsurprisingly respond to an ambitious reform by surrendering more of the contested resource. In contrast, when reform is centrally implemented, citizens give credit for improvements on the status quo, regardless of whether those are actually fulfilled locally (see Equation 2); this reduces elites’ incentive to implement reform.

Equation 5 illustrates a central fact about bargaining models with an outside option when the incentive to “rebel” (start a war, etc.) is driven by expressive rather than instru-
mental motivations: the grievances of the “dissatisfied” party are incompletely internalized by the “satisfied” party. To see this, it is useful to compare our model to a “conventional” take-it-or-leave-it bargaining model with an outside option, where the dissatisfied party is instrumentally motivated to obtain the contested resource. In such a model, a higher probability of successful rebellion has two effects—it increases the incentive of the dissatisfied party to rebel, and it increases the cost of rebellion to the satisfied party—which together imply that the satisfied party fully internalizes any change in that probability.\textsuperscript{10} In contrast, in a bargaining model where the incentive to rebel is driven by expressive motivations, a change in some parameter that affects grievances (here, $\gamma$) in general affects the incentive to rebel only.

Our central question is the relationship between rebellion and reform, which is given by

$$\frac{\partial F(\Gamma(x^*; \gamma))}{\partial \gamma} = f(\Gamma(x^*; \gamma)) \left[ (\lambda - \beta) - \lambda \cdot \frac{\partial x^*}{\partial \gamma} \right].$$

The direction of this relationship is equal to the sign of the expression in brackets, which corresponds to the direct effect of reform on grievances, holding local implementation constant, and an indirect effect through the elite’s equilibrium choice of $x$. Using Equation 5, this expression can be rewritten for an interior solution as

$$[1 - \alpha(x^*)] (\lambda - \beta).$$

(6)

The relationship between rebellion and reform is thus proportional to the direct effect of

\textsuperscript{10}That is, consider a model in which the satisfied party solves

$$\max_x -x + F(p - x) \cdot (x - p),$$

where $p$ is the probability of successful rebellion. Then

$$\frac{\partial x^*}{\partial p} = 1 + \frac{H'(p - x^*)}{[H(p - x^*)]^2} \left( 1 - \frac{\partial x^*}{\partial p} \right) = 1$$

. 

13
reform on rebellion \((\lambda - \beta)\), holding local implementation constant, with a discount that depends on the curvature of the hazard rate \(H\).

Local implementation of reform drives a wedge between citizens’ expectations of what reform promises and what actually happens. Elites may attempt to minimize this wedge, and thus the risk of rebellion, by carrying out some of what reform demands, but they discount the grievances arising from under-implementation. At the same time, the citizenry “gives credit” (where the parameter \(\beta\) measures the degree of credit) for reforms that improve on the status quo—this reduces grievances and the propensity to rebel. The former effect dominates the latter when reform is predominantly locally implemented.

We are now in a position to offer some preliminary answers to the questions raised in the introduction:

- **To what extent do the “certain groups” in a position to block implementation of reform internalize the resulting rebellion?** To a partial extent only. The expressive grievances in our model are only incompletely internalized by local elites.
- **Is there always a tradeoff between stability and reform?** Yes. Any reform that incentivizes local implementation necessarily also increases the risk of rebellion. Further below we explore whether there exists a tradeoff between aggregate (local and central) implementation and stability.
- **Is reform indeed most dangerous for “bad governments”?** If by “bad government” we mean one with limited capacity to centrally implement reform, then yes. Reform is indeed most dangerous for such governments.

## 2 Robustness

The simple model of the previous section generates some sharp results. It is reasonable to ask whether these are driven by a number of simplifying assumptions. In this section we demonstrate the robustness of our results to various extensions and generalizations. (Where
appropriate, we provide detailed proofs in the online appendix.) We begin by relaxing the assumption of a unitary citizenry.

### 2.1 Coordination within localities

In the previous section, we analyze a model in which the probability that the citizenry rebels is \( F(\Gamma (x; \gamma)) \), where \( F \) is the distribution of the random cost of rebellion \( \epsilon \) and

\[
\Gamma (x; \gamma) = \lambda (\gamma - x) - \beta (\gamma - \hat{\gamma}).
\]

Here we derive this relationship from a simple model of strategic interaction among local citizens, whose coordinated participation is necessary for rebellion to be successful, and a leader who decides whether to allow rebellion to proceed.

There is a continuum of citizens, indexed by \( j \). Each citizen decides whether to participate in a local rebellion. Any citizen who chooses not to rebel receives a payoff normalized to zero. Consider now the payoff from participation. Any citizen \( j \) has grievance

\[
\tilde{\Gamma} (x; \gamma) \equiv \Gamma(x; \gamma) + \xi,
\]

where the parameter \( \xi \geq \beta (\gamma - \hat{\gamma}) \). Following Passarelli and Tabellini (2017), we assume that the expressive payoff from participating in rebellion is proportional to the grievance and the (endogenous) proportion \( g \) of citizens who rebel.\(^{11}\) Any citizen who participates further bears a cost \( \epsilon + u_j \). We treat \( \epsilon \) as exogenous—that is, as realized prior to the decision of individual citizens to participate (but after the choice of \( x \) by the local elite). The idiosyncratic component \( u_j \), in turn, is drawn from a uniform distribution on \([-\phi, \phi]\).

\(^{11}\)Together with the assumption below that rebellion is more likely to proceed when joined by many citizens, this formalization recalls Wood’s (2003) notion of “pleasure in agency,” in which the psychological benefits from rebellion are related to the probability of success. See Morris and Shadmehr (2018) for discussion.
With these assumptions, the marginal benefit of participation is

$$g \tilde{\Gamma} (x; \gamma) - \varepsilon - u_j.$$

To guarantee a unique level of rebellion in equilibrium (see Lemma A1 and the surrounding discussion in the online appendix for an analogous argument when strategic complementarities are across rather than within localities), we assume that there is sufficient heterogeneity in the idiosyncratic cost of rebellion:

$$\tilde{\Gamma}(0; \gamma) < 2\phi. \quad (8)$$

Depending on its distribution, $\varepsilon$ (realized before citizens decide whether to participate) may take values sufficiently low or high such that all or no citizens, respectively, choose to rebel. For moderate values of $\varepsilon$, the proportion of citizens who participate is given implicitly by

$$g = g \cdot \frac{\tilde{\Gamma} (x; \gamma) - \varepsilon + \phi}{2\phi},$$

which follows from the assumption that $u_j$ is distributed uniformly on $[-\phi, \phi]$. Solving for $g$ gives

$$g = \phi - \varepsilon \frac{2\phi - \tilde{\Gamma} (x; \gamma)}{2\phi}.$$

Condition 8 guarantees that the denominator of this expression is positive for all $x \in [0, \gamma]$ and thus that there is a unique level of rebellion in equilibrium.

To derive $F(\Gamma (x; \gamma))$ as defined above, we additionally assume that there is a local citizen leader who allows rebellion to proceed if and only if $g \geq \bar{g}$, where $\bar{g}$ is an exogenous threshold that determines that rebellion is “worth it.” Intuitively, we can think of the citizen leader as being in a position to apply selective incentives not to participate if it appears that rebellion is unlikely to be successful. From the perspective of the local elite, which treats $\varepsilon$ as a random variable, the probability of rebellion is therefore

$$\Pr \left( \frac{\phi - \varepsilon}{2\phi - \tilde{\Gamma} (x; \gamma)} \geq \bar{g} \right) = \Pr \left( \frac{\varepsilon}{\bar{g}} - \xi + \frac{\phi}{\bar{g}} (2\bar{g} - 1) \leq \Gamma (x; \gamma) \right),$$
where the equality uses the definition of $\tilde{\Gamma}(x; \gamma)$ in Equation 7. Defining $\epsilon \equiv \epsilon/\bar{g} - \xi + \frac{\bar{g}}{\bar{g}} (2\bar{g} - 1)$ gives $F(\Gamma(x; \gamma))$.

Although not completely satisfactory—the assumption of a local citizen leader is ad hoc, though perhaps reasonable—the formalization here illustrates that it is possible to microfound the baseline model in a theory of local collective action. We now proceed to consider various generalizations of that model.

2.2 Instrumental motivations to rebel

In the model above, we assume that the citizenry is motivated by expressive concerns only. As discussed, this reasonably captures an environment in which material gains from rebellion are non-excludable. Nonetheless, it is straightforward to incorporate instrumental motivations, such that the citizenry is motivated to rebel not only for expressive reasons but also to seize control of the local resource.

Consider in particular the following extension of the baseline model. As before, we assume that, in the event of rebellion, the elite loses (in expectation) proportion $p \in (0, \gamma)$ of the resource under local control. We now additionally assume that the citizenry gains this same proportion. In contrast, if no rebellion is attempted, the elite transfers $\lambda x$ to the citizenry. Finally, regardless of whether rebellion is attempted or successful, the elite transfers the exogenously mandated share, valued at $(1 - \lambda) \gamma$. The expected material gain to the citizenry from rebellion is therefore $\lambda(p - x)$. We parameterize the relative importance of this gain by $\theta$. When citizens are motivated by expressive concerns only—for example, because the material gains from revolution are non-excludable—then $\theta = 0$, corresponding to the baseline model.

With these assumptions, the elite’s problem is the same as before. The motivation to rebel, however, is different: the citizenry’s marginal payoff from rebellion is now

$$\theta \lambda(p - x) + \Gamma(x; \gamma) - \epsilon.$$
Given these assumptions, at an interior solution, the equilibrium level of local implementation is given implicitly by
\[ x^* = p - \frac{1}{(1 + \theta)\lambda} \cdot \left( 1 - \frac{1 - F(z(x^*; \gamma))}{f(z(x^*; \gamma))} \right), \]
where
\[ z(x; \gamma) \equiv \theta \lambda (p - x) + \Gamma(x; \gamma). \]
From this it follows that the equilibrium relationship between local implementation and reform at an interior solution can be expressed as
\[ \frac{\partial x^*}{\partial \gamma} = \hat{\alpha}(x^*) \cdot \frac{\lambda - \beta}{(1 + \theta)\lambda}, \quad (9) \]
where
\[ \hat{\alpha}(x^*) \equiv \left[ 1 + \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2} \right]^{-1} \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2}. \]

In this extended model, the relationship between rebellion and reform is given by
\[ \frac{\partial F(z(x^*; \gamma))}{\partial \gamma} = f(z(x^*; \gamma)) \left[ (\lambda - \beta) - (1 + \theta)\lambda \cdot \frac{\partial x^*}{\partial \gamma} \right]. \]
Using Equation 9, the expression in brackets can be rewritten for an interior solution as
\[ [1 - \hat{\alpha}(x^*)] (\lambda - \beta), \]
which is directly analogous to Expression 6 in the baseline model. Notably, the sign of the relationship between rebellion and reform is independent of the degree \( \theta \) to which the citizenry values the material payoff from rebellion, though in general the magnitude of the relationship depends on \( \theta \) (through \( f \) and \( H \)).

### 2.3 Coordination across localities

Finally, we examine robustness to relaxing the assumption that local rebellions are isolated one from another. In practice, unrest often spills over into neighboring areas, strengthening the resolve of citizens there to act upon their grievances.
To explore the impact of coordination across localities, consider the following extension of the baseline model. There is a continuum of localities, indexed by \( i \). In each locality \( i \), the local elite chooses implementation \( x_i \), following which the local citizenry chooses to rebel. Similarly to Passarelli and Tabellini (2017), we assume that the expressive payoff from rebellion is proportional to the grievance and to the participation of citizens in other localities. In particular, let the net payoff from rebellion be

\[
h \tilde{\Gamma}(x_i; \gamma) - \epsilon_i,
\]

where \( h \in [0, 1] \) is the endogenous proportion of localities that rebel, and the grievance

\[
\tilde{\Gamma}(x_i; \gamma) \equiv \lambda (\gamma - x_i) - \beta (\gamma - \hat{\gamma}) + \xi.
\]

The random cost of rebellion \( \epsilon_i \) is drawn independently across localities from a common distribution \( F \) with strictly increasing hazard rate. We assume the parameter \( \xi \geq \beta (\gamma - \hat{\gamma}) \), which implies that the expressive payoff from rebellion is positive so long other localities also rebel (e.g., because of the “warm glow” from participating in collective action), even if \( x_i = \gamma \). Then if there is sufficient heterogeneity in the idiosyncratic cost of rebellion \( \epsilon_i \) (an assumption analogous to “two-sided limit dominance” in the literature on global games, given assumed preferences; see Morris and Shin, 2003), for any common level of local implementation \( \bar{x} \in [0, \gamma] \) there is a unique level of rebellion \( h^*(\bar{x}) \).

We use this result to derive a symmetric equilibrium, in which the elite in any locality \( i \) chooses \( x_i = \bar{x} \). Notwithstanding the different strategic environment, the equilibrium level of local implementation takes a form similar to that in Equation 4:

\[
\bar{x} = \max \left[ p - \frac{1}{h^*(\bar{x}) \lambda} \cdot \frac{1 - F \left( h^*(\bar{x}) \tilde{\Gamma}(\bar{x}; \gamma) \right)}{f \left( h^*(\bar{x}) \tilde{\Gamma}(\bar{x}; \gamma) \right)}, 0 \right].
\]

From this it follows that the relationship between local implementation and reform can be expressed as

\[
\frac{\partial \bar{x}}{\partial \gamma} = \tilde{\alpha}(\bar{x}) \cdot \frac{\lambda - \beta}{\lambda},
\]
where $\tilde{\alpha}(\tilde{x})$ is a fraction bounded by zero and one. The relationship between reform and rebellion is proportional to the same expression. Our key comparative statics thus extend to the case in which elites and citizens take into account the interdependence of rebellion across localities.

In summary, the equilibrium behavior in our baseline model is robust to various generalizations and extensions. We now turn to implications of that behavior.

3 Implications

As shown above, whether reform provokes or prevents rebellion depends on whether local agents with a stake in the status quo are primarily responsible for its implementation. What determines whether this is the case? We see two possibilities.

First, when state capacity is relatively high, it may be possible to implement reform centrally. Capacious states are characterized by comparatively large bureaucracies (Schiavo-Campo, de Tommaso and Mukherjee, 1997; Gehlbach, 2008; Brown, Earle and Gehlbach, 2009) organized around Weberian principles of meritocratic recruitment and internal promotion (Rauch and Evans, 2000). These features endow bureaucracies with the ability to implement reform as intended (Huber and McCarty, 2004; Ting, 2011) and in a manner congruent with local conditions (Bendor and Meirowitz, 2004). Absent this capacity, central governments may be forced to delegate authority to local agents with little interest in reform for its own sake. In some cases, the resulting “drift” from government intent can be minimized through the monitoring of these agents, which essentially amounts to central implementation, but weak states are often generally weak: the same limits to state capacity that necessitate local implementation may prevent the central government from holding local elites accountable.

Second, some reforms by their nature are easier than others to implement centrally.

\footnote{\textsuperscript{12}For a review and further discussion, see Gailmard and Patty (2012).}
Universal rights can often be mandated from above, though even here monitoring may be necessary to ensure that those rights are enforced. Programs of redistribution, in contrast, may involve the reallocation of local property rights, which are often poorly understood by central authorities.

Our theory also has implications for the optimal choice of reform, though we do not model the selection of $\gamma$ directly. Imagine a reformist but skittish central government interested in effecting the transfer of some contested resource from local elites to the citizenry but concerned about accompanying unrest. What is the optimal reform design?

On the one hand, regardless of the degree $\lambda$ of local control over the implementation of reform, aggregate (local plus central) reform implementation is greater when reform is more ambitious. To see this, observe that aggregate reform implementation is

$$\lambda x^* + (1 - \lambda) \gamma,$$

where $x^*$ is equilibrium local implementation. Differentiating with respect to $\gamma$ gives

$$\lambda \frac{\partial x^*}{\partial \gamma} + (1 - \lambda),$$

which is clearly positive if $x^* = 0$. Focus, then, on the case $x^* > 0$, and consider the extended model that allows for instrumental motivations. (The analysis is nearly identical for the model with coordination across localities.) Recalling that, for this case, $\frac{\partial x^*}{\partial \gamma} = \hat{\alpha} (x^*) \cdot \frac{\lambda - \beta}{\lambda(1 + \theta)}$, where $\hat{\alpha} (x^*)$ is a fraction bounded by zero and one, we can rewrite Expression 11 as

$$\hat{\alpha} (x^*) \cdot \frac{\lambda - \beta}{1 + \theta} + (1 - \lambda),$$

which is clearly positive for $\lambda \geq \beta$. For the residual case in which $x^* > 0$ but $\lambda < \beta$, observe that

$$\frac{1 - \lambda}{\beta - \lambda} > 1 > \frac{\hat{\alpha} (x^*)}{1 + \theta},$$

given that $\theta \geq 0$.

On the other hand, when the nature of reform or limits to state capacity necessitate local control over implementation (i.e., when $\lambda > \beta$), a more ambitious reform also implies more
rebellion, as shown in various versions of our model. At some level, this simply reflects the reality that in many environments the threat of rebellion is the only way to incentivize local elites to make concessions. Nonetheless, rebellion typically comes at a cost. When it does, the optimal reform design may set $\gamma < 1$, as the central government’s interest in reform is balanced against its desire for stability.

4 Empirical illustration: Russia’s emancipation of the serfs

In this section we illustrate the insights gained from the model by analyzing an important case of institutional reform: Russia’s emancipation of the serfs in 1861. As we demonstrate, local implementation in a context of weak state capacity provoked unrest among precisely those that reform was designed to help.

4.1 Background

In Russia, serfdom developed differently from elsewhere in Europe. Indeed, through the fifteenth century, Russian peasants enjoyed legal freedom and the right to choose their place of residence and employment. The gradual introduction of serfdom was driven by two factors. First, state building and territorial expansion necessitated the creation of a large class of military and civilian servitors, who were compensated for their service by land grants (e.g., Kimerling Wirtschafter, 2008). Second, however, territorial expansion into unsettled territories, coupled with peasants’ freedom of movement, put the economic wellbeing of these servitors at risk (Domar, 1970). Ever-increasing restrictions on peasants’ rights made state service more appealing for servitors, and these restrictions ultimately culminated in the formal introduction of serfdom in the mid-seventeenth century.

Russian serfs were legally “attached” to the land on which they lived, and the estate’s aristocratic owner enjoyed almost unlimited powers over her serfs. By law, serfs were not al-
owed to own property, and the estate’s owner enjoyed broad policing and judicial powers over the serfs, including the authority to administer various punishments. The majority of serf peasants were allotted a strip of land for cultivation and subsistence, in exchange for which they were required to perform certain obligations; the most significant of these were unpaid labor on the landlord’s fields (*barshchina*, or corvée), payment in money or kind (*obrok*, or quitrent), or some combination of the two. Serfs were at constant risk of being stripped of their land, which belonged entirely to the landlord, resettled, or sold. Although in theory there were some limitations on landowners’ behavior, such as the restriction of *barshchina* to no more than three days a week, such safeguards were tenuous at best, as the law prescribed corporal punishment for any complaints against an estate’s owners (Zaionchkovskii, 1968).

The logic behind serfdom was the Russian state’s model of universal service: serfs served and provided income for the nobles, who in turn were legally obligated to serve the state. Over time, however, the nobility’s obligations under this social contract were eliminated. Empress Anna (1730–40) “allowed nobles to keep at least one son at home to run the estate” (Montefiore, 2016, p. 156), and in 1762 the service requirement for the nobility was abolished altogether. With this change, the moral justification for serfdom was also eliminated.

For Russia’s rulers, serfdom presented a political as well as moral problem. As in any system of coerced labor, there were periodic spasms of violence, from the murder of individual landlords to large-scale peasant revolts that threatened the very survival of the monarchy, including the Pugachev Rebellion (1773–1775). At the same time, any attempts to reform the institution were blocked by the nobility, on whose support the Romanov dynasty relied and for whom the existing economic order was quite profitable (Domar and Machina, 1984). Tsar Paul I (1796–1801), who attempted to limit the nobility’s power over their serfs and to introduce minimum land allotments, was murdered in a palace coup (Zenkovsky, 1961, p. 282). “Serfdom is a powder magazine under the state and the peasantry is an explosive mine,” admitted Tsar Nicholas I (1825-55), yet “to tamper with it now would be, of course, an even more disastrous evil” (Volin, 1943, p. 48). Any attempts at reform were restricted
to the periphery of the Empire, such as the landless emancipation of the Baltic serfs between 1816 and 1819, or were extremely limited in scope, such as the introduction in the 1840s of “inventories” meant to regulate serfs’ obligations and land allotments in right-bank Ukraine.

The government’s attitude changed with Russia’s humiliating defeat in the Crimean War (1853–1856), which exposed the country’s backwardness and its inability to compete in the international order. The most important of the “Great Reforms” that followed was Russia’s emancipation of the serfs. In 1857, Tsar Alexander II (1855–1881) warned Moscow’s nobility that it would be better to emancipate the serfs “from above” than to allow this to happen “from below.” On December 4, 1858, Alexander publicly announced that serfdom would be abolished, regardless of the fears and desires of the noble class.

While the intent to emancipate the serfs was widely publicized, the exact content of the reform was subject to negotiations and pitched political battles at court. A large portion of the nobility viewed emancipation as an existential threat to their economic wellbeing and lobbied for landless emancipation. The so-called “liberal bureaucracy,” on the other hand, pushed for the distribution of land among emancipated former serfs. The emancipation act of 1861, publicized as the “Manifesto of February 19,” was a political compromise that left no group fully satisfied. Serfs were granted immediate personal freedom and the right to own personal property. The reform also introduced minimum and maximum personal land allotments, which varied with soil type. The landowner and peasants had the option to agree on an immediate “grant allotment” of one-quarter of the maximum allotment, for which the peasant would not be required to pay or provide obligations. If the peasants wished instead to receive their full land allotment, they became “temporarily obligated” until such time as the landowner chose to begin the “redemption” operation that transferred ownership to the peasantry. The terms of temporary obligation—and, typically, the subsequent redemption process—were to be established during a two-year transition period, as regulatory charters (ustavnye gramoty) specifying land allotments, payments, and the general framework of landlord-peasant relations were compiled by the landlord in cooperation with the peasantry.
The Russian government was fully aware that the reform’s content, while improving serfs’ lot by providing legal freedom and the distribution of some land, nevertheless did not represent the maximum possible improvement on the status quo. At the same time, reformers expected some slippage between emancipation’s promise and its reality, as we discuss below. Preparations were thus made for disturbances following publication of the Manifesto, with stallions ready to spirit the tsar from a rebelling capital, should that prove necessary (Zaionchkovskii, 1968, p. 159).

As anticipated, emancipation did indeed provoke substantial unrest. As Finkel, Gehlbach and Olsen (2015) show, there was a substantial increase in disturbances among former serfs during the transition period described above, with no corresponding change among state and appanage peasants (peasants who lived on state lands and land owned by the imperial family, respectively), who were not directly affected by the reform. Moreover, much of the unrest occurred not immediately after publication of the Manifesto, but the following year, as reform was being implemented on individual estates. Across the empire, thousands of peasants refused to provide *barshchina* or *obrok*, physically attacked landowners and their families, took up arms, and engaged in other forms of resistance.

What accounted for the peasant unrest that followed emancipation? Our model suggests that reform provokes rebellion when state capacity is weak and grievances are driven by the failure of local actors to fully implement reform. We address each of these factors below.

### 4.2 State capacity, local implementation, and perceptions of loss

A key feature of the emancipation reform was the inability of the central government to carry out reform on its own. Outside of major urban centers, the Russian state’s control of its territory was limited at best. As Skocpol (1979) notes, “[i]mperial jurisdiction stopped just outside the doors of the noble-owned serf estates” (p. 89), which is precisely where the government needed to be to implement the reform. The state’s weakness was an inevitable outcome of the monarchy’s conscious policy of reliance on the nobility as its local agents.
Ironically, the peasant revolts of the seventeenth and eighteenth centuries had “convince[d] the state of the value of the nobility as a police force in the provinces” (Moon, 2001, p. 27).

The government’s weakness extended to the very center of the state apparatus. The Russian government did not have the institutional capacity or the manpower to carry out reform on its own. It didn’t even know the lay of the land. Throughout the 1840–50s, the Russian government carried out a number of cadastral surveys, but no national cadastre existed (Evtuhov, 2011). The land surveyors who could have been tasked with carrying out a national cadastre were in short supply and of uncertain professional abilities (Khristoforov, 2011; Dower and Markevich, 2017). The political and security police (the Third Section of the Imperial Chancellery and the Special Corps of Gendarmes, respectively) had fewer than five thousand servicemen for the entire empire. The only way to carry out the reform was by delegating its implementation to local actors with a stake in the pre-existing status quo.

Reform’s local implementation became its key challenge. In principle, the peasants were entitled to their existing land allotments, but so ambitious a reform and the very fact that the reform’s content was a compromise among numerous conflicting policy proposals provided local elites with ample opportunities for gerrymandering. Fearful for their economic wellbeing in a new, post-serfdom world, numerous landlords jumped at the opportunity to keep the estate’s best land for themselves (especially in regions with fertile soil) and to ensure that former serfs received as little valuable land as possible.

Anticipating potential conflicts between dissatisfied peasants and landlords, as part of the emancipation reform the government created the new institution of “peace arbitrators” (мировые посредники), tasked with the verification of charters and the resolution of conflicts between landlords and the newly liberated peasants. Nonetheless, even for this institution the weak Russian state was forced to rely on local elites. Regional governors were asked to find reform sympathizers from among the local landowning (and often serf-owning) nobility to fill these positions; Leo Tolstoy was a notable example (Ust’iantseva, 1992; Easley, 2002). In some areas, however, supporters of emancipation among the nobility were nowhere to be
found, and government efforts notwithstanding, individuals of “every political stripe, with varying degrees of vulnerability to local pressures” were drafted into the institution (Easley, 2002, p. 711). Many found it hard to be neutral while the interests of neighboring landlords were at stake, and some used outright violence, including flogging, to compel peasants to accept the charter terms.

Even when the arbitrators were willing to confront local elites, the landlords were often able to neutralize the “troublesome” mediators by using a combination of political, psychological, and even physical pressure. “They want to beat me up, they want me to be put on trial . . . I am simply waiting until they calm down a bit (pougomonilis’) and then I will retire,” wrote Tolstoy about his relations with local landlords (Ust’iantseva, 1992, p. 179). Tolstoy’s experience is far from being unique: from 1861 to 1863, more than 25 percent of arbitrators quit their roles, often as a result of pressure and hostility from landowners (Easley, 2002, p. 727). With peace arbitrators sidelined and with no accountability to the central government, the implementation of reform became a purely local process.

Against this backdrop, it is not surprising that former serfs blamed the local nobility for their failure to fully implement reform. The design of reform played to this tendency. Most serfs learned of emancipation when the Manifesto was read out in Sunday church services. This public articulation of the tsar’s intention may have set a new (and common) reference point against which the implementation of reform was measured. The actual outcome of reform would thus have been substantially experienced as a loss, relative to that with which peasants were “endowed” by the reform’s formal design.

With peasant grievances directed toward local elites with a stake in the status quo, a weak state could not hope to pacify the peasantry. The price of reform was rebellion.
5 Conclusion

Tocqueville’s conjecture was that citizens would be most likely to rebel precisely when reform promised to change their lives for the better. We provide a microfoundation for this claim, showing that reforms that are introduced centrally but implemented locally create the conditions for rebellion by driving a wedge between what citizens expect and what they receive. By setting a reference point against which local implementation falls short, reform creates grievances that may be only partially offset by reform’s improvement over the status quo. It is the failure of officials to fully internalize these grievances that produces the Tocqueville paradox.

At the same time, our analysis places bounds on the effect that Tocqueville describes. It is only when reform is “complicated” or state capacity is low that reform provokes rebellion. When central governments are able to push through implementation on their own, reform unambiguously reduces rebellion.

Our results imply a need to recalibrate theories of regime change and autocratic stability. Even ignoring intertemporal commitment problems of the sort emphasized by Acemoglu and Robinson (2006), the ability of governments to buy off excluded groups should not be taken for granted. When the complexity of reform overwhelms the capacity of the state to carry it out, the implementation of reform will typically fall short of its promise, producing feelings of loss that encourage rebellion against those responsible. Future work can productively extend existing models to incorporate this perspective.

References


Easley, Roxanne. 2002. “Opening Public Space: The Peace Arbitrator and Rural Politiciza-


University Press.


Skocpol, Theda. 1979. States and Social Revolutions: A Comparative Analysis of France, Russia, and China. Cambridge: Cambridge University Press.


Ting, Michael M. 2011. “Organizational Capacity.” Journal of Law, Economics, and Or-


The Tocqueville Paradox: Additional Material

A generalized model

This section formally expresses the generalized model with instrumental as well as expressive motivations. The baseline model corresponds to the special case in which $\theta = 0$.

The following assumptions ensure that the random variable $\epsilon$ is strictly increasing on an interval sufficiently wide to ensure a non-degenerate probability of rebellion.

**Assumption A1.** $F (\theta \lambda (p - \gamma) - \beta (\gamma - \hat{\gamma})) > 0$.

**Assumption A2.** $F (\lambda (\theta p + \gamma) - \beta (\gamma - \hat{\gamma})) < 1$.

We additionally assume

**Assumption A3.** For all $s \in S$, where $S$ is the support of $F$, the hazard rate $H (s) \equiv \frac{f(s)}{1-F(s)}$ is strictly increasing.

Define

$$z (x; \gamma) \equiv \theta \lambda (p - x) + \Gamma (x; \gamma) \quad (A1)$$

$$= \theta \lambda (p - x) + \lambda (\gamma - x) - \beta (\gamma - \hat{\gamma}).$$

The next three propositions immediately follow.

**Proposition A1.** Assume that Assumptions A1–A3 are satisfied. Then the optimal local implementation of reform $x^*$ is given implicitly by

$$x^* = \max \left[ p - \frac{1}{(1 + \theta) \lambda} \cdot \frac{1 - F (z (x^*; \gamma))}{f (z (x^*; \gamma))}, 0 \right]. \quad (A2)$$

**Proof.** The solution follows directly from maximization of Problem 3, given the generalized probability of rebellion assumed here. To see that a strictly increasing hazard rate ensures that the second-order condition is satisfied for an interior solution, observe that the second derivative of Equation 3 with respect to $x$ is

$$-f (z (x; \gamma)) (1 + \theta) \lambda - [f (z; \gamma (x)) - (x - p) f' (z (x; \gamma)) (1 + \theta) \lambda] (1 + \theta) \lambda,$$

A1 (intended for online publication)
which is less than zero in a neighborhood of \( x = x^* \) if
\[
\frac{f'(z(x^*; \gamma))}{f(z(x^*; \gamma))} > -\frac{2}{(p-x^*)(1+\theta)\lambda}.
\]
(Observe from Equation A2 that \( p > x^* \).) Substituting from Equation A2 gives
\[
\frac{f'(z(x^*; \gamma))}{f(z(x^*; \gamma))} > -2 \cdot \frac{f(z(x^*; \gamma))}{1-F(z(x^*; \gamma))}.
\]
(A3)

Now observe that a strictly increasing hazard rate implies, for all \( z \),
\[
\frac{f'(z)[1-F(z)] + f(z)^2}{[1-F(z)]^2} > 0,
\]
which in turn implies
\[
\frac{f'(z)}{f(z)} > -\frac{f(z)}{1-F(z)}.
\]
(A4)

Clearly, if Condition A4 holds for all \( z \), then Condition A3 must also hold.

Proposition A2. Assume that Assumptions A1–A3 are satisfied. Then a marginal increase in reform \( \gamma \) increases (decreases) local reform implementation if and only if \( \lambda \) is greater (less) than \( \beta \).

Proof. Differentiating Equation A2 implicitly with respect to \( \gamma \) for the case \( x^* > 0 \) gives
\[
\frac{\partial x^*}{\partial \gamma} = \frac{1}{(1+\theta)\lambda} \cdot \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2} \left[ (\lambda - \beta) - (1+\theta)\lambda \cdot \frac{\partial x^*}{\partial \gamma} \right],
\]
where \( H' \) is the derivative of \( H \). Rearranging gives
\[
\frac{\partial x^*}{\partial \gamma} = \left[ 1 + \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2} \right]^{-1} \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2} \cdot \frac{\lambda - \beta}{(1+\theta)\lambda}.
\]
(A5)

where
\[
\hat{\alpha}(x^*) \equiv \left[ 1 + \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2} \right]^{-1} \frac{H'(z(x^*; \gamma))}{[H(z(x^*; \gamma))]^2}.
\]
As \( H'(z(x^*; \gamma)) > 0 \) by assumption of a strictly increasing hazard rate, \( \frac{\partial x^*}{\partial \gamma} \geq 0 \) if and only if \( \lambda \geq \beta \).

\[ \square \]
Proposition A3. Assume that Assumptions A1–A3 are satisfied. Then the equilibrium probability of rebellion is increasing (decreasing) in reform $\gamma$ if and only if $\lambda$ is greater (less) than $\beta$.

Proof. For $x^* = 0$, the statement follows directly from Equation A1 and the assumption that $F$ is strictly increasing. Now assume $x^* > 0$. Differentiating the probability of rebellion with respect to $\gamma$ gives

$$
\frac{\partial F(z(x^*;\gamma))}{\partial \gamma} = f(z(x^*;\gamma)) \left[ (\lambda - \beta) - (1 + \theta) \lambda \cdot \frac{\partial x^*}{\partial \gamma} \right].
$$

Substituting from Equation A5 gives

$$
\frac{\partial F(z(x^*;\gamma))}{\partial \gamma} = f(z(x^*;\gamma)) \cdot \left[ 1 - \hat{\alpha}(x^*) \right] (\lambda - \beta),
$$

the sign of which is given by $(\lambda - \beta)$.

Coordination across localities

The analysis to follow assumes for verisimilitude that citizens observe local implementation only in their own locality, but this is without loss of generality: the analysis is similar, and our results identical, if the choice of $x_i$ for each locality $i$ is commonly observed. Letting $I$ denote the set of all localities, we define an equilibrium as follows.

Definition A1. An equilibrium of the model with spillovers is a strategy profile

$$
\beta = \left( (x_i)_{i \in I}, (\gamma_i(e_i))_{i \in I} \right),
$$

and a belief assessment $\mu$, such that

1. For all $i$, $x_i = \bar{x}$.

2. The pair $(\beta, \mu)$ is sequentially rational and consistent.

Condition 1 restricts attention to equilibria that are symmetric, in the sense that elites choose a common level of reform implementation. Condition 2 is simply the definition of sequential equilibrium applied to our setting.

A3 (intended for online publication)
The assumption of symmetric equilibrium implies that the “rebellion game” (i.e., the subform that follows choice of $x_i$ in each locality $i$) takes a particular form. Consider the decision to rebel in any locality $i$. The citizenry in $i$ chooses to rebel if

$$\epsilon_i < h\tilde{\Gamma}(x_i; \gamma).$$

(A6)

In principle, localities may differ not only in their idiosyncratic cost of rebellion $\epsilon_i$, but also in their beliefs about the level of reform implementation in localities other than their own, which by assumption is unobserved. Condition 1 of the equilibrium definition, however, requires that elites in all localities $i$ choose a common $x_i = \bar{x}$. Moreover, the consistency requirement of sequential equilibrium implies that citizens believe that $\bar{x}$ has been chosen in all localities other than their own. Together, these considerations imply that the proportion of localities that rebel is defined implicitly by

$$h = F\left(h\tilde{\Gamma}(\bar{x}; \gamma)\right).$$

(A7)

It is straightforward to establish that Equation A7 has a unique fixed point for all $\bar{x} \in [0, \gamma]$ if there is sufficient heterogeneity in the idiosyncratic cost of rebellion $\epsilon_i$.

**Assumption A4.** $F(0) > 0$.

**Assumption A5.** $F\left(\tilde{\Gamma}(0; \gamma)\right) < 1$.

**Lemma A1.** [Passarelli and Tabellini, 2017] Assume that Assumptions A4 and A5 are satisfied. For all $\bar{x} \in [0, \gamma]$, there exists a fixed point $h^*(\bar{x}) \in (0, 1)$ of Equation A7. This fixed point is unique if

$$\tilde{\Gamma}(\bar{x}; \gamma) \cdot f\left(h^*(\bar{x})\tilde{\Gamma}(\bar{x}; \gamma)\right) < 1.$$  

(A8)

**Proof.** Define

$$d(h) = F\left(h\tilde{\Gamma}(\bar{x}; \gamma)\right) - h.$$ 

By Assumption A4, for all $\bar{x} \in [0, \gamma]$, there are localities that would choose to rebel if no others do so, implying $d(0) > 0$. Similarly, by Assumption A5, $d(1) < 0$ for all $\bar{x} \in [0, \gamma]$, as A4 (intended for online publication)
there are localities that would choose not to rebel even if all others do so. Together, these two conditions imply that any fixed point of Equation A7 is interior. Existence of such a fixed point is guaranteed by Brouwer’s fixed-point theorem, given the continuity (because differentiable) of $F$. Uniqueness, in turn, is established by Condition A8, which implies that $d(h)$ is decreasing in $h$ in the neighborhood of any solution.

In what follows, we use the following assumption, which ensures that Condition A8 holds for all $\bar{x} \in [0, \gamma]$.

Assumption A6. For all $\bar{x} \in [0, \gamma]$, $\Gamma (\bar{x}; \gamma) \cdot f \left( h^* (\bar{x}) \Gamma (\bar{x}; \gamma) \right) < 1$.

Substituting $h^* (\bar{x})$ into Condition A6 gives the probability that the citizenry in locality $i$ rebels as

$$F \left( h^* (\bar{x}) \Gamma (x_i; \gamma) \right).$$

(A9)

The elite in locality $i$ anticipates the relationship between reform implementation and rebellion captured by this expression in solving

$$\max_{x_i} -x_i + F \left( h^* (\bar{x}) \Gamma (x_i; \gamma) \right) (x_i - p).$$

(A10)

Taking the first-order condition and imposing $x_i = \bar{x}$ gives the equilibrium level of local implementation.

Proposition A4. Assume that Assumptions A3–A6 are satisfied. Then the common local implementation of reform $\bar{x}$ in the model with coordination across localities is given implicitly by

$$\bar{x} = \max \left[ \frac{p - 1}{h^* (\bar{x}) \lambda}, \frac{1 - F \left( h^* (\bar{x}) \Gamma (\bar{x}; \gamma) \right)}{f \left( h^* (\bar{x}) \Gamma (\bar{x}; \gamma) \right)}, 0 \right].$$

(A11)

Proof. Equation A11 follows directly from Problem A10. A proof analogous to that for Proposition A1 ensures that the second-order condition is satisfied for an interior solution.

The next two propositions immediately follow.

A5 (intended for online publication)
**Proposition A5.** Assume that Assumptions A3–A6 are satisfied. If $\bar{x} > 0$, a marginal increase in reform $\gamma$ increases (decreases) local reform implementation in the model with coordination across localities if and only if $\lambda$ is greater (less) than $\beta$.

**Proof.** Differentiating Equation A11 implicitly with respect to $\gamma$ for the case $\bar{x} > 0$ gives

$$\frac{\partial \bar{x}}{\partial \gamma} = \frac{1}{\lambda} \left[ \frac{h'}{H^2} \cdot \frac{1}{H} \cdot \frac{\partial \bar{x}}{\partial \gamma} + \frac{H'}{H^2} \cdot \frac{1}{h} \left( h' \Gamma (\bar{x}; \gamma) \frac{\partial \bar{x}}{\partial \gamma} + h \left( \lambda - \lambda \frac{\partial \bar{x}}{\partial \gamma} - \beta \right) \right) \right], \tag{A12}$$

where $h'$ is the derivative of $h$ with respect to $\bar{x}$; we drop arguments for notational compactness. Rearranging gives

$$\frac{\partial \bar{x}}{\partial \gamma} = \left[ 1 - \frac{1}{\lambda} \frac{H'}{H^2} \cdot \frac{1}{h} \left( \frac{h'}{h^2} \cdot \frac{1}{H} \cdot \frac{\partial \bar{x}}{\partial \gamma} + \frac{H'}{H^2} \cdot h' \Gamma (\bar{x}; \gamma) \right) + \frac{H^2}{H} \frac{\lambda - \beta}{\lambda} \right]^{-1} \frac{H'}{H^2} \cdot \frac{1}{h} \left( h' \Gamma (\bar{x}; \gamma) \frac{\partial \bar{x}}{\partial \gamma} + h \left( \lambda - \lambda \frac{\partial \bar{x}}{\partial \gamma} - \beta \right) \right).$$

By assumption of a strictly increasing hazard rate, $H' > 0$. Further, differentiating Equation A7 implicitly with respect to $\bar{x}$ and rearranging gives

$$\frac{\partial h}{\partial \bar{x}} = -\frac{h' \lambda f}{1 - \Gamma (\bar{x}; \gamma) f} < 0.$$

Thus, $\frac{\partial \bar{x}}{\partial \gamma} \geq 0$ if and only if $\lambda \geq \beta$. \qed

**Proposition A6.** Assume that Assumptions A3–A6 are satisfied. The equilibrium probability of rebellion in any locality $i$ in the model with coordination across localities is increasing (decreasing) in reform $\gamma$ if and only if $\lambda$ is greater (less) than $\beta$.

**Proof.** Differentiating Expression A9 with respect to $\gamma$ gives

$$f \cdot \left[ h' \Gamma (\bar{x}; \gamma) \frac{\partial \bar{x}}{\partial \gamma} + h \left( \lambda - \beta - \lambda \frac{\partial \bar{x}}{\partial \gamma} \right) \right].$$

For $\bar{x} = 0$, $\frac{\partial \bar{x}}{\partial \gamma} = 0$ and the statement in the proposition immediately follows. For $\bar{x} > 0$, we can use Equation A12 to substitute for the equation in brackets, giving

$$f \cdot \frac{\partial \bar{x}}{\partial \gamma} \left( 1 - \frac{1}{\lambda} \frac{h'}{h^2} \cdot \frac{1}{H} \right) \frac{H^2}{H'}.$$

As $H' > 0$ (by assumption of a strictly increasing hazard rate) and $h' < 0$ (see the proof to Proposition A5), the equilibrium probability of rebellion is increasing (decreasing) in $\gamma$ if and only if $\bar{x}$ is increasing (decreasing) in $\gamma$, and thus (by Proposition A5) if and only if $\lambda$ is greater (less) than $\beta$. \qed