POLITICAL CENTRALIZATION AND GOVERNMENT ACCOUNTABILITY*

FEDERICO BOFFA
AMEDEO PIOLATTO
GIACOMO A. M. PONZETTO

This article explains why decentralization can undermine accountability and answers three questions: what determines if power should be centralized or decentralized when regions are heterogeneous? How many levels of government should there be? How should state borders be drawn? We develop a model of political agency in which voters differ in their ability to monitor rent-seeking politicians. We find that rent extraction is a decreasing and convex function of the share of informed voters, because voter information improves monitoring but also reduces the appeal of holding office. As a result, information heterogeneity pushes toward centralization to reduce rent extraction. Taste heterogeneity pulls instead toward decentralization to match local preferences. Our model thus implies that optimal borders should cluster by tastes but ensure diversity of information. We also find economies of scope in accountability that explain why multiplying government tiers harms efficiency. A single government in charge of many policies has better incentives than many special-purpose governments splitting its budget and responsibilities. Hence, a federal system is desirable only if information varies enough across regions. *JEL* Codes: D72, D82, H73, H77.

*For helpful comments we are grateful to the editor and three anonymous referees; to Peter Backus, Christopher Berry, Paula Bustos, Vasco Carvalho, Steve Cicala, Philippe De Donder, Ruben Enikolopov, Gino Gancia, Edward Glaeser, Matthew Kahn, Yuriy Kaniovskyi, Gianmarco León, Ana Nuevo-Chiquero, Luigi Pascali, Maria Petrova, Diego Puga, Thijs van Rens, Kurt Schmidheiny, Hannes Schwandt, Paolo Surico, Jaume Ventura and Yanos Zylberberg; to seminar participants at Barcelona, Bolzano, Bristol, CEMFI, Chicago Harris, Collegio Carlo Alberto, CREI, EIEF, Erasmus University Rotterdam, Florence, IESE, IMT Lucca, LUISS, Nottingham, Pompeu Fabra, Toulouse, Vanderbilt, and Yale; and to conference participants at APET, c.MET05, IIOC, NBER Summer Institute, RES, SED, Urban Economic Association, and VATT. Boffa acknowledges financial support from the Free University of Bolzano (WW82); Piolatto from the Spanish Ministry of Science and Innovation (ECO-2012-37131) and the Government of Catalonia (2014-SGR-420); Ponzetto from the Spanish Ministry of Science and Innovation (JCI-2010-08414 and ECO-2011-25624), the Spanish Ministry of Economy and Competitiveness (RYC-2013-13838), the Government of Catalonia (2009-SGR-1157 and 2014-SGR-830), the Barcelona GSE and the BBVA Foundation through its first grant for Researchers, Innovators and Cultural Creators. The opinions expressed in this project belong to the authors, and the BBVA Foundation is not responsible for them.

The Quarterly Journal of Economics (2016), 381–422. doi:10.1093/qje/qjv035.
Advance Access publication on September 28, 2015.

I. INTRODUCTION

In the run-up to Scotland's 2014 independence referendum, the Scottish government published a guide setting out its case for independence. Alex Salmond, the premier, argued that Scotland ought to become independent because its people are different from those of other parts of the British Isles and thus need a different government of their own. "After Scotland becomes independent . . . the people of Scotland are in charge. It will no longer be possible for governments to be elected and pursue policies against the wishes of the Scottish people" (Salmond 2013, pp. x–xi).

The Scottish leader's argument finds support in the standard economic theory of fiscal federalism. Its core result is the decentralization theorem: absent policy spillovers, decentralization is more efficient than centralization if regions are not identical. This proposition, introduced by Oates (1972), has proved a remarkably general paradigm (Lockwood 2006). Local governments can tailor their choices to the particular conditions of each jurisdiction and thus provide higher social welfare than a single policy adopted by a common government. With no economies of scale, each group with distinct preferences should have an independent government (Tiebout 1956; Bewley 1981). Increasing returns and externalities promote political integration, but heterogeneity raises the downsides of large jurisdictions (Alesina and Spolaore 2003). Political-economy frictions provide rigorous microfoundations for the inability of a central government to match local preferences (Lockwood 2002; Besley and Coate 2003; Harstad 2007).

Yet empirical evidence shows that decentralization has not consistently delivered the benefits its advocates predicted in theory (Treisman 2007). The majority of Scottish voters who rejected independence in the referendum may have been risk averse but not unwise. The experience of countries all over the world teaches that decentralization can harm the quality of government just as it can improve it. Mismanagement and lack of accountability are common in local governments, especially in developing and transition economies (Bardhan and Mookherjee 2006).

This article develops a model of political agency that explains why decentralization can reduce accountability and answers three key questions. When regions are heterogeneous, what determines if power should be centralized or decentralized? How

many levels of government should there be? How should state borders be drawn? Our theory is grounded on the observation that regions differ not only in preferences—the focus of the classic theory—but also in their ability to monitor elected officials and hold government accountable. Government accountability varies widely within the United States: official corruption in Louisiana and Mississippi is five times as prevalent as in Oregon and Washington (Glaeser and Saks 2006).

We study public goods provided by self-interested politicians whose goal is to extract wasteful rents. To keep extracting rents they need to win reelection, so their corruption is constrained by career concerns. Electoral discipline provides both incentives and screening. Politicians differ in ability and voters try to dismiss unskilled incumbents. Voters infer skill from performance, so politicians are incentivized to refrain from extracting rents because low public-good provision is punished at the polls, whether it stems from incompetence or corruption.

Our model has two key features. First, we study heterogeneous accountability arising from differences in voters' information. Some voters correctly observe and understand policy outcomes, whereas others do not and cannot infer the incumbent's ability. Second, we develop a dynamic model with a recursive incentive structure. The expectation of future electoral discipline affects the current trade-off between rent extraction and reelection. Thus, a permanent increase in voter information has two effects on electoral discipline. On the one hand, it makes reelection more responsive to performance, raising incentives to reduce rents. On the other hand, this very reduction in equilibrium rents lowers the appeal of reelection and thus indirectly dampens the decline in rent extraction. In our model we find that the direct effect always dominates, but rent extraction falls with voter information at a declining rate because of the countervailing indirect effect. When monitoring improves starting from a low initial level, politicians react sharply because the value of office is high. Further improvements yield lower benefits.

Our core theoretical insight follows from the concave impact of an informed population on the quality of government. When different regions have different shares of informed voters, centralization reduces aggregate rent extraction. Political integration creates a single electorate with the average share of informed voters. Rent extraction falls sharply in less informed

regions, while it does not increase as much in better informed ones. Thus, centralization yields aggregate efficiency gains.

However, the distribution of these efficiency gains is problematic. A centralized government is more accountable, but disproportionately accountable to the most informed regions. If it enjoys discretion over the geographic distribution of public goods, if favors informed regions and neglects uninformed ones. The resulting misallocation is regressive and so costly that centralization lowers social welfare despite reducing rents. Thus, we find that centralization can be welfare-maximizing only if it is accompanied by a uniformity constraint that requires the central government to provide identical public goods to all regions.

As a result of this endogenous need for uniformity, heterogeneous information drives a key trade-off. Centralization improves accountability, but it forgoes the ability to match local public goods to idiosyncratic local preferences. Section III analyzes this trade-off and answers our motivating question: should government be decentralized when regions are different? The answer depends on what type of heterogeneity is starkest. Differences in tastes pull toward decentralization; differences in information push toward centralization instead.

Empirical evidence supports our results. Without a uniformity constraint, politicians allocate spending across regions in response to voter information rather than actual needs (Strömberg 2004). With uniformity, instead, centralization mainly benefits the uninformed: reforms decentralizing public education in Argentina and Italy had regressive effects and worsened inequality (Galiani, Gertler, and Schargrodsky 2008; Durante, Labartino, and Perotti 2014).

Our prediction that centralization improves government accountability is consistent with U.S. history. Two former state governors—Don Siegelman of Alabama and Rod Blagojevich of Illinois—are in prison for corruption. Corruption has long been considered a distinctive plague of city and state governments (Steffens 1904; Wilson 1966). Federal intervention during the New Deal eradicated the patronage and political manipulation that had characterized state and local welfare programs until then (Wallis, Fishback, and Kantor 2006). World history offers other examples of accountability gains from centralization: in early modern Europe (Besley and Persson 2011; Dincecco 2011), in precolonial Africa (Gennaioli and Rainer 2007), and in transition economies (Blanchard and Shleifer 2001).

European history also provides direct support for our finding that heterogeneous accountability prompts centralization. Germany and Italy were unified as nation-states in the late nineteenth century. Italy had highly heterogeneous preunitary institutions and became a centralized nation-state. Instead, Germany had relatively homogeneous institutional quality and was organized as a federal country. Both regional differences in accountability and the degree of centralization remain higher in Italy than Germany today (Ziblatt 2006).

In Section IV we study how many levels of government there should be. The standard logic of fiscal federalism suggests there should be many because every policy should be matched to the right geographic unit. In our framework, however, we find that multiplying government tiers is costly because there are economies of scope in accountability. When politicians are responsible for providing a larger set of public goods, their incentives improve and they devote a lower share of the budget to rents. Such economies of scope imply that having a single level of government is best if information is homogeneous. A federal system can be optimal only if differences in information are large enough. Then the federal government provides a large accountability gain to poorly informed regions, while their local governments can match their idiosyncratic preferences over policies for which taste heterogeneity is starkest.

Our model can thus explain the empirical finding that government quality declines as the number of government tiers rises. In the United States, the proliferation of overlapping special-purpose local governments in charge of specific policies has been a fiasco (Berry 2009). Special-purpose districts are inefficient and prone to capture by special interests. In Europe, too, multiple subnational levels of governments have led to inefficiencies, and their reduction and simplification is now on the agenda. Cross-country evidence shows a robust positive correlation between corruption and the number of levels of government (Fan, Lin, and Treisman 2009).

Section V considers what should determine the boundaries of governments when people are not naturally sorted into internally homogeneous regions. We find that optimal borders have two characteristics: they cluster by tastes but ensure maximum diversity of information. The second goal can trump the first when geographic constraints create a tension between the two. A disadvantaged uninformed group should not be a local minority; it

should rather join better informed voters with similar preferences in a larger polity. For example, breaking up California would reduce welfare because educated San Francisco liberals ought to share a state government with working-class left-wingers in the Central Valley.

This article furthers the study of fiscal federalism and the geographic structure of government. Starting with Tiebout's (1956) and Oates's (1972) seminal contributions, prior work focused exclusively on differences in preferences. We show that this is only half of the story. Once we consider differences in voter information across regions, we find that the two kinds of heterogeneity have opposite implications on the optimal architecture of government.

Differences in preferences promote decentralization if the central government cannot tailor policies to local preferences (Oates 1972; Alesina and Spolaore 2003). Assuming that accountability is homogeneous across regions, prior work endogenized the failure of preference matching under centralization through frictions in political bargaining (Lockwood 2002; Besley and Coate 2003; Harstad 2007). We provide an alternative microfoundation through heterogeneous voter information.

More important, we show that differences in information promote centralization because they entail larger accountability gains from political integration. Our finding suggests that heterogeneous information is the key reason centralization can increase accountability. Prior work mainly emphasized why accountability can rise with decentralization. In particular, decentralization can help voters monitor their local governments thanks to yardstick competition (Besley and Case 1995), while centralization entails a common agency problem that makes politicians less accountable to voters in any single region (Seabright 1996).[1]

Furthermore, we provide the first theory of economies of scope in government accountability. Prior work considered each policy instrument in isolation, typically assessing if it would be best centralized or decentralized (Oates 1999). We extend this line of inquiry by studying the pros and cons of a federal structure

1. Although we are not the first to model accountability gains from centralization, the potential sources of such gains with homogeneously informed voters have always proven theoretically ambiguous (Lockwood 2006; Treisman 2007). In our framework, instead, centralization unambiguously alleviates moral hazard in political agency.

with multiple levels of government in charge of providing distinct public goods.[2]

II. Political Agency and Public-Good Provision

In this section, we present the model of political agency that underpins our analysis of optimal political integration. Imperfectly informed voters face the problem of selecting and incentivizing self-interested rent-seeking politicians. We model electoral discipline in a framework of political career concerns (Alesina and Tabellini 2008). Voters try to retain competent politicians and dismiss incompetent ones. In solving this screening problem, they also create incentives for politicians to provide public goods. The incumbent moderates rent extraction because higher public-good provision raises voters' inference of his ability and thus his chances of reelection.

II.A. Preferences and Technology

The economy is populated by a continuum of infinitely lived agents, whose preferences are separable over time and additive in utility from private consumption and utility from each of P public goods. Individual i in period t derives instantaneous utility

$$u_t^i = \tilde{u}_t^i + \sum_{p=1}^{P} \alpha_p^i \log g_{p,t}, \tag{1}$$

where $\tilde{u}_t^i$ is exogenous utility from private consumption, and $g_{p,t}$ the provision of public good p. We treat $\tilde{u}_t^i$ as an exogenous mean-zero shock and focus exclusively on public goods. Each public good yields benefits according to a logarithmic utility function. The relative importance of each good for individual i is described by the ideal shares $\alpha_p^i \geq 0$ such that $\sum_{p=1}^{P} \alpha_p^i = 1$.

Each public good p is produced by the government with technology

$$g_{p,t} = e^{\eta_{p,t}} x_{p,t}. \tag{2}$$

The production technology has constant returns to scale: $x_{p,t}$ measures per capita investment in providing public good p. We rule out economies of scale in public-good provision, which would provide an immediate technological rationale for centralization.

2. Online Appendix A provides a more complete discussion of the literature.

Productivity $\eta_{p,t}$ represents the stochastic competence of the incumbent politician in providing public good p. It follows a first-order moving average process:

$$\eta_{p,t} = \varepsilon_{p,t} + \varepsilon_{p,t-1}. \tag{3}$$

The shocks $\varepsilon_{p,t}$ are independent and identically distributed across goods, over time, and across politicians. They have support $[\check{\varepsilon}, \hat{\varepsilon}]$, mean zero, and variance σ^2. Our preferred interpretation is that parties are composed of overlapping generations of politicians. The period t government consists of older party leaders with competence $\varepsilon_{p,t-1}$ and young party members with competence $\varepsilon_{p,t}$. At $t+1$, former party leaders retire, rising young politicians take over the leadership, and a new cohort joins the party.

Politicians are self-interested rent seekers. Their objective is to maximize the present value of the rents they can extract while in office, discounted by the discount factor $\delta \in (0, 1]$. Each period, the government allocates a fixed government budget b. The incumbent chooses the amount $x_{p,t}$ of expenditure on each public good. He extracts as rent the remainder,

$$r_t = b - \sum_{p=1}^{P} x_{p,t}, \tag{4}$$

which represents public resources devoted to socially unproductive projects.[3]

II.B. Elections and Information

The incumbent faces reelection at the end of each period. If ousted he will never return to power. Politicians cannot make policy commitments, so the election is not based on campaign promises but on retrospective evaluation of the incumbent's track record. Voters do not directly observe the incumbent's competence or his actions. Their inference is based on an imperfect signal of public-good provision. The textbook model of career concerns assumes that voters observe policy outcomes with additive noise. We assume instead that voter information is binary. An

3. Rent extraction could identically be interpreted as slacking (Seabright 1996; Alesina and Tabellini 2008). Politicians enjoy an "ego rent" b from holding office. However, they incur a cost $x_{p,t}$ from exerting effort to provide public goods. Then r_t then captures politicians' failure to work diligently in their constituents' interest.

informed voter observes perfectly the vector $\mathbf{g}_t$ of realized public goods. An uninformed voter receives no informative signal of $\mathbf{g}_t$, or proves completely incapable of understanding information about $\mathbf{g}_t$.[4]

The electorate consists of a continuum of atomistic voters, partitioned into J groups. Group j comprises a fraction λ_j of voters, whose preferences are described by the vector $\boldsymbol{\alpha}^j$ of their ideal shares. The share of group j members who are informed about public-good provision is a random variable Θ_t^j, independent and identically distributed over time. Our model is robust to an arbitrary correlation of information across voters.[5] We measure voter information by the expected share of informed voters $\theta_j = \mathbb{E}\Theta_t^j$.

We allow for an intensive margin of political support, following the probabilistic voting approach. Each voter's preferences consist of two independent elements. First, agents have preferences over the public goods they expect either politician (the incumbent I or the challenger C) to provide in the following period. These preferences are summarized by the difference

$$\Delta^i \equiv \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \left(\log g_{p,t+1}^I - \log g_{p,t+1}^C \right), \tag{5}$$

where $\mathbb{E}_i$ denotes the rational expectation given voter i's information. Second, voters have preferences for candidates' characteristics other than their competence: for example, personal likability or party ideology. These preferences can be decomposed into an aggregate shock Ψ_t and an idiosyncratic shock ψ_t^i that is independent and identically distributed across voters.

4. Uninformed voters may not realize that public goods affect their utility. Such ignorance is particularly natural for public goods that yield long-run benefits. Voters may also understand the benefits of public goods, but fail to understand how they depend on the incumbent's actions and competence (Strömberg 2004).

5. Most simply, information could be uncorrelated across voters. Each voter in group j has probability θ_j of being informed. Then in every period a share θ_j of group members are informed. This assumption is consistent with imperfect sharing of information within a group (Ponzetto 2011; Ponzetto and Troiano 2014). First, agents privately acquire information. Some fail to observe $\mathbf{g}_t$. Second, agents communicate with a finite number of neighbors. Some remain uninformed because none of their neighbors observed $\mathbf{g}_t$. If instead information sharing is perfect, information is perfectly correlated within each group. With probability θ_j the entire group is informed ($\Theta_t^j = 1$), and with probability $1 - \theta_j$ the entire group is uninformed ($\Theta_t^j = 0$).

Voting is costless, and all voters cast a ballot for their preferred candidate. Thus, voter i votes for the incumbent if and only if $\Delta^i \geq \Psi_t + \psi_t^i$. As in Baron (1994) and Grossman and Helpman (1996), informed voters cast their ballot based on observed policy outcomes, while uninformed voters choose which candidate to support purely on the basis of preferences unrelated to competence.[6]

The distribution of the shocks Ψ_t and ψ_t^i is symmetric around 0, so voters do not systematically favor incumbents or challengers. We assume that the two shocks are uniformly distributed: $\Psi_t \sim U[-\frac{1}{2\phi}, \frac{1}{2\phi}]$ and $\psi_t^i \sim U[-\overline{\psi}, \overline{\psi}]$. The support of preference shocks is wide enough and the support of competence innovations $\varepsilon_{p,t}$ narrow enough that

$$\frac{1}{2\phi} - \overline{\psi} \leq \check{\varepsilon} < \hat{\varepsilon} \leq \overline{\psi} - \frac{1}{2\phi} \text{ and } -\frac{1}{2\phi} \leq \check{\varepsilon} < \hat{\varepsilon} \leq \frac{1}{2\phi}. \tag{6}$$

The first set of inequalities ensures that every voter's ballot is imperfectly predictable, irrespective of $\mathbf{g}_t$. The second ensures that the outcome of the election is never entirely predictable either.

The timeline within each period t is the following.

(i) The incumbent's past competence innovations $\boldsymbol{\varepsilon}_{t-1}$ become common knowledge.
(ii) The incumbent chooses investments $\mathbf{x}_t$ and rent r_t.
(iii) The competence innovations $\boldsymbol{\varepsilon}_t$ are realized and the provision of public goods $\mathbf{g}_t$ is determined.
(iv) Voter information is realized: a share Θ_t^j of members of group j perfectly observe $\mathbf{g}_t$. The rest remain completely uninformed. No voter has any direct observation of $\boldsymbol{\varepsilon}_t$, $\mathbf{x}_t$, or r_t.
(v) An election is held, pitting the incumbent against a single challenger, randomly drawn from the same pool of potential officeholders.

II.C. Political Career Concerns

Voters rationally expect every politician to choose the stationary investment $\overline{\mathbf{x}}$. The equilibrium allocation is time-

6. The standard assumption that uninformed voters vote sincerely could be attributed to their imperfect rationality (Baron 1994; Grossman and Helpman 1996). It is also consistent with full strategic rationality because a continuum of voters entails strategic insignificance: no voter can ever be pivotal.

invariant because the environment is stationary. It does not vary with the incumbent's observed skills ε_{t-1} because performance is separable in effort and ability. It cannot vary with the competence innovations ε_t because they are unknown to the politicians themselves when they make policy choices.[7] Thus, the outcome of the election affects future public-good provision only through differences in politicians' skills:

$$(7)\quad \Delta^i = \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \left(\eta_{p,t+1} - \eta_{p,t+1}^C \right) = \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \left(\varepsilon_{p,t} - \varepsilon_{p,t}^C \right) = \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \varepsilon_{p,t}.$$

No information exists about future competence innovations (either the incumbent's ε_{t+1} or the challenger's ε_{t+1}^C) or about the challenger's current ability (ε_t^C). Thus, their expectation is nil for all voters. Uninformed voters cannot infer the incumbent's ability from realized public-good provision and retain the unconditional expectation $\mathbb{E}\varepsilon_{p,t} = 0$.[8] Informed voters, instead, can infer the incumbent's ability from their knowledge of public-good provision:

$$(8)\qquad \mathbb{E}(\varepsilon_{p,t} | g_{p,t}) = \log g_{p,t} - \log \overline{x}_p - \varepsilon_{p,t-1}.$$

In a rational expectation equilibrium their inference is perfectly accurate ($x_{p,t} = \overline{x}_p$ entails $\mathbb{E}(\varepsilon_{p,t} | g_{p,t}) = \varepsilon_{p,t}$).

From the politician's perspective, the probability of reelection as a function of his policy choices is

$$(9)\qquad \pi(\mathbf{x}_t) = \frac{1}{2} + \phi \sum_{j=1}^{J} \theta_j \lambda_j \sum_{p=1}^{P} \alpha_p^j (\log x_{p,t} - \log \overline{x}_p),$$

7. The agent's lack of private information is the defining technical feature of career concern models. A more complicated signaling model in which politicians privately observe their own ability before choosing their costly hidden action delivers the same qualitative results on incentives and screening in the political agency problem (Banks and Sundaram 1998).

8. We assume that uninformed voters vote sincerely based on their unconditional expectation because they are strategically insignificant or imperfectly rational (Baron 1994; Grossman and Helpman 1996). With a finite number of voters, an uninformed voter with full strategic rationality would instead care about his vote only when it is pivotal. In the limit as the number of voters diverges, uninformed voters would vote strategically based on expected ability conditional on an exactly tied election. In the equilibrium of our model, this conditional expectation remains $\mathbb{E}_i \varepsilon_{p,t} = 0$ given that the aggregate taste shock Ψ_t is uniformly distributed on a sufficiently large support. Thus, we could identically assume that uninformed voters have a pivotal voting motivation provided they cannot infer the aggregate taste shock Ψ_t from their own tastes $\Psi_t + \psi_t^i$, either because the idiosyncratic shock is diffuse ($\overline{\psi} \to \infty$) or because their Bayesian reasoning is imperfect.

as we derive in Online Appendix C. The incumbent faces a trade-off. Investing in public goods reduces his rents but increases his chances of reelection by raising informed voters' inference of his ability. A politician who values reelection R chooses to extract rents

$$r = b - \phi R \sum_{j=1}^{J} \theta_j \lambda_j. \tag{10}$$

In a dynamic equilibrium, the value of reelection R is the expected present value of future rents from holding office. In a rational expectation equilibrium voters cannot be fooled ($\overline{x}_p = x_{p,t}$). Then in every election the incumbent wins with probability $\pi = \frac{1}{2}$. Voter preferences are not exogenously biased in favor of incumbents or against them (the distribution of Ψ_t and ψ_t^j is symmetric around zero). An endogenous incumbency advantage does not arise because politicians' ability evolves as a first-order moving average process.[9] As a consequence, a politician who rationally anticipates extracting rent r whenever in office has an expected net present value of reelection

$$R = \delta \sum_{t=0}^{\infty} \left(\frac{\delta}{2}\right)^t r = \frac{2\delta}{2-\delta} r. \tag{11}$$

II.D. Government Accountability from Voter Information

Let $\rho \equiv \frac{r}{b} \in [0, 1]$ denote the fraction of the budget allocated to rents. The unique stationary rational expectations equilibrium has the following characterization.[10]

LEMMA 1. In equilibrium, ruling politicians extract rents

$$\rho = \left(1 + \frac{2\delta}{2-\delta} \phi \sum_{j=1}^{J} \theta_j \lambda_j\right)^{-1}$$

9. The impact of each competence shock lasts for two periods only, so past screening of incumbents does not translate in a forward-looking electoral advantage as it does with longer-lasting competence shocks (Ashworth and Bueno de Mesquita 2008). If the period-t incumbent was reelected at $t-1$ the expectation of current productivity $\boldsymbol{\eta}_t$ is above average. Senior party leaders have proved their competence and won reelection. However, their known ability $\boldsymbol{\varepsilon}_{t-1}$ is orthogonal to future performance $\boldsymbol{\eta}_{t+1}$ because they are about to retire. A new cohort leads the party into the period t election. Their skills $\boldsymbol{\varepsilon}_t$ can be inferred from policy outcomes $\mathbf{g}_t$, but not from the past reelection of their retiring colleagues.

10. All proofs are provided in Online Appendix C.

and have expected ability

$$\mathbb{E}\hat{\eta}_{p,t} = \phi\sigma^2 \sum_{j=1}^{J} \alpha_p^j \theta_j \lambda_j.$$

Rent extraction is a decreasing and convex function of voter information ($\frac{\partial \rho}{\partial \theta_j} < 0$ and $\frac{\partial^2 \rho}{\partial \theta_j^2} > 0$). An increase in voter information θ_j increases the ability of ruling politicians $\hat{\eta}_{p,t}$ in the sense of first-order stochastic dominance.

Better information improves government accountability because it enables voters to monitor politicians more closely. It alleviates both the moral hazard problem of politicians' incentives and the adverse selection problem of politicians' selection. Voters can reward public-good provision only when they perceive it accurately. As voter knowledge improves, the incumbent's performance more closely determines his chances of reelection. Ex ante, he extracts lower rents because his career concerns are heightened ($\frac{\partial \rho}{\partial \theta_j} < 0$). Ex post, the average ability of ruling politicians rises because electoral screening improves ($\frac{\partial \mathbb{E}\hat{\eta}_{p,t}}{\partial \theta_j} > 0$).[11]

The key result in Lemma 1 is that rent extraction is decreasing but convex in voter information ($\frac{\partial^2 \rho}{\partial \theta_j^2} > 0$).[12] Decreasing returns to monitoring follow from the dynamic nature of the politicians' problem. The direct impact of voter information on rent extraction is linear (equation (10). For a given value of reelection R, more informed voters induce one-to-one more investment and lower political rents. A transitory one-period increase in voter information would have no other effect, but a permanent increase in voter information has an indirect effect, too. Politicians

11. Voters have no incentives to acquire information to improve governance because of the rational voter paradox. Each voter has a negligible likelihood of determining the outcome of the election. His strategic incentives to become informed are likewise negligible. Therefore, information θ_j reflects exogenous voter characteristics. For example, education enables voters to grasp the precise role of politicians in providing public goods; social capital reflects civic involvement and the ability to share political knowledge in a wide social network (Ponzetto and Troiano 2014).

12. Other determinants of the quality of government are straightforward. More patient politicians are more willing to reduce rent extraction to raise their chances of reelection ($\frac{\partial \rho}{\partial \delta} < 0$). A higher variance of politicians' ability raises the gains from screening ($\frac{\partial \mathbb{E}\hat{\eta}_{p,t}}{\partial \sigma^2} > 0$). Both incentives and screening improve when voters are keener on competence than other determinants of political popularity ($\frac{\partial \rho}{\partial \phi} < 0$ and $\frac{\partial \mathbb{E}\hat{\eta}_{p,t}}{\partial \phi} > 0$).

understand and expect tighter monitoring if they are reelected, so the expected future rents from holding office decrease. Their decline reduces the incentives to refrain from extracting rents and mitigates the direct effect of improved monitoring. Current rent extraction is more sensitive to the expectation of future rents when voters' average information is higher. Thus, a marginal increase in voters' information causes a lower decline in rent extraction when the share of informed voters is higher to begin with.[13]

A large body of evidence confirms that the quality of government is higher if citizens are more educated and politicians are subject to greater media scrutiny (e.g., Glaeser et al. 2004; Svensson 2005; Glaeser and Saks 2006; Snyder and Strömberg 2010). While none of these studies have explored specifically the concavity of this relationship, the data provide suggestive empirical support for our prediction. Svensson (2005) documents that low human capital is the best predictor of high corruption across countries. Consistent with Lemma 1, Figure I shows that corruption is not only a decreasing but also a convex function of the share of people with a tertiary education. A similar relationship emerges in Figure II, where we proxy information with newspaper circulation instead. Both results are robust to controlling for income.[14]

Our finding that government accountability is an increasing but concave function of voter information has a broader theoretical underpinning. The mechanism in Lemma 1 applies to any

13. Extreme cases highlight decreasing returns to monitoring with particular clarity. If no voters are informed, career concerns are absent and rent extraction is unchecked ($\theta = 0 \Rightarrow \rho = 1$). Introducing a little monitoring induces a forceful reaction by politicians who are afraid of losing very large rents. If all voters are informed, career concerns are at their strongest but rent extraction cannot be reduced to 0 ($\theta = 1 \Rightarrow \rho > 0$). Incumbents always extract some rents: only the appeal of future rents induces them to make any productive investment. Marginally worsening perfect monitoring causes a small loss.

14. The multivariate regressions are respectively $\rho_l = \underset{(.5)}{2.4} - \underset{(.06)}{.23} ln\, y_l - \underset{(5)}{26}\theta_l + \underset{(27)}{82}\theta_l^2 + \varepsilon_l$ for education (across 118 countries) and $\rho_l = \underset{(.5)}{1.6} - \underset{(.07)}{.11} ln\, y_l - \underset{(2)}{12}\theta_l + \underset{(3)}{13}\theta_l^2 + \varepsilon_l$ for newspaper circulation (across 100 countries). Corruption $\rho_l \in [-2.5, 2.5]$ is the opposite of the Control of Corruption index from the World Governance Indicators, averaging across available years (1996–2013). Real GDP per capita is from the Penn World Tables 8.0, measured in 1970 following Svensson (2005). The share of people over age 25 with a tertiary education is from Barro and Lee (2013), also measured in 1970. Newspaper circulation per capita is from the World Development Indicators, averaging across available years (1997–2005).

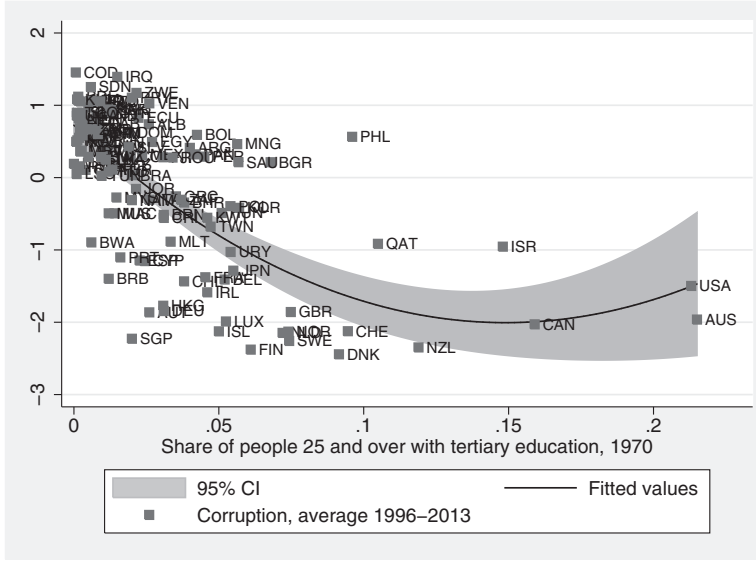

FIGURE I

Corruption and Education

Corruption is the opposite of the Control of Corruption index from the World Governance Indicators. The share of people over age 25 with tertiary education is from Barro and Lee (2013).

determinant of electoral discipline. Information, however, has an additional source of concavity: it can be shared by voters. The share Θ_t^j of informed voters then results from a two-stage process (Ponzetto 2011; Ponzetto and Troiano 2014). First, it includes those who acquired information directly, for instance, because they read newspapers or because their human capital enables them to assess politicians' performance accurately. Second, it includes those who did not acquire information directly but obtained it from an informed neighbor. Overall, the expected share of informed voters θ_j is an increasing and concave function of the probability that each voter acquires information directly, because one voter's knowledge has greater spillovers if his neighbors are less informed.[15]

15. If each agent obtains information directly with probability $\underline{\theta}_j$ and shares it in a group of n neighbors, his eventual probability of being informed is $\theta_j = 1 - \left(1 - \underline{\theta}_j\right)^n$ such that $\frac{\partial \theta_j}{\partial \underline{\theta}_j} > 0 > \frac{\partial^2 \theta_j}{\partial \underline{\theta}_j^2}$.

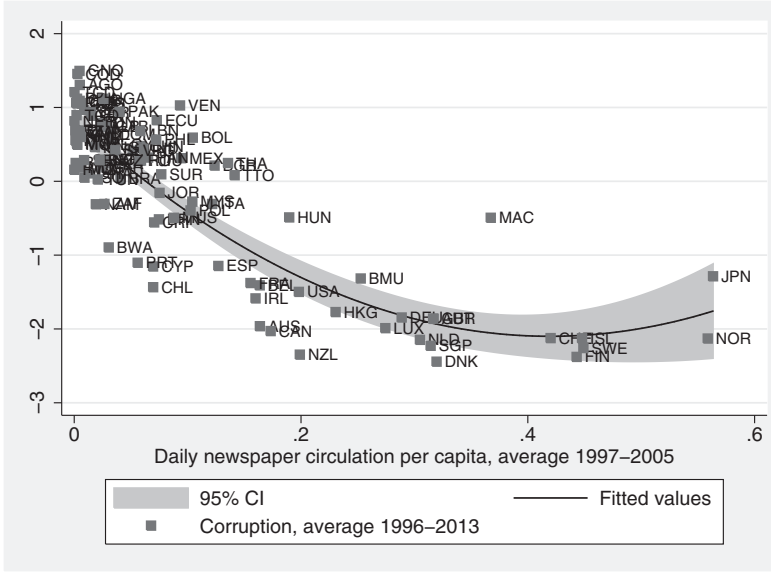

FIGURE II

Corruption and Newspaper Circulation

Corruption is the opposite of the Control of Corruption index from the World Governance Indicators. Newspaper circulation per capita is from the World Development Indicators.

III. SHOULD GOVERNMENT BE DECENTRALIZED?

We turn now to our motivating question. Should different regions have different governments whenever there are no spillovers, in accordance with Oates's (1972) classic decentralization theorem? When can we expect decentralization to deliver the benefits Salmond touted to Scotland's voters? When will centralization curb the graft and mismanagement of local governments, as with welfare spending and the New Deal (Wallis, Fishback, and Kantor 2006)? The key to our answer is that regions differ along several dimensions. They have different preferences but also different levels of voter information.

We consider an economy composed of L regions, each populated by a unit measure of voters. Preferences are homogeneous within each region, but heterogeneous across regions (Tiebout 1956; Oates 1972). For example, residents of conservative "red states" may prefer greater spending on defense, justice, and

police, whereas residents of progressive "blue states" may prefer instead environmental protection, public education, and welfare spending. Our novel contribution lies in studying at the same time differences in voter information. For example, states with more educated residents have a higher expected share of informed voters, and voters in less educated states are less likely to assess government performance accurately.

Formally, we assume that each region's preference vector $\boldsymbol{\alpha}^l$ is an independent draw from a distribution that is symmetric across goods, so the marginal distribution α_p^l is the same for all p and has mean $\mathbb{E}\alpha_p^l = \frac{1}{P}$.[16] Then preference heterogeneity can range between two limit cases. It is nil when $\alpha_p^l = \frac{1}{P}$ deterministically. In this limit case of perfectly homogeneous preferences, everyone desires the same uniform basket of public goods. At the opposite extreme, preference heterogeneity is maximized when α_p^l has a Bernoulli distribution with $\Pr(\alpha_p^l = 1) = \frac{1}{P}$. In this limit case of maximum preference heterogeneity, each region desires a single idiosyncratic public good, so the same good yields utility to two regions with negligible probability $\frac{1}{P}$. A series of mean-preserving spreads gradually spreads out the distribution of preferences from the first limit case to the second. We parameterize the distribution of preferences by a homogeneity parameter $\upsilon \in \mathbb{R}^+$ such that the distribution becomes less dispersed as υ increases, spanning the whole feasible range. That is, an increase in υ entails a mean-preserving contraction of α_p^l. The limit case of maximum preference heterogeneity corresponds to $\upsilon = 0$ and the limit case of perfectly homogeneous preferences to $\upsilon \to \infty$.[17]

Information is independent of preferences. Each region's expected share of informed voters θ_l is an independent draw from a distribution with mean $\mathbb{E}\theta_l = \overline{\theta} \in (0, 1)$. Then, information heterogeneity can range between two limit cases. It is nil when $\theta_l = \overline{\theta}$ deterministically. In this limit case of perfectly homogeneous information, every region has the same expected share of informed voters. At the opposite extreme, information heterogeneity is maximized when θ_l has a Bernoulli distribution with $\Pr(\theta_l = 1) = \overline{\theta}$. In this limit case of maximum information heterogeneity, a fraction $\overline{\theta}$

16. We abstract from differences between the sample distribution and the population distribution by considering the limit case of a continuum of regions.

17. For example, $\boldsymbol{\alpha}^l$ could have a symmetric Dirichlet distribution on the regular $(P-1)$-simplex with concentration parameter $\upsilon > 0$. Our results do not rely on this particular specification.

of regions are perfectly informed ($\theta_l = 1$) while the remainder $1 - \overline{\theta}$ are completely uninformed ($\theta_l = 0$). A series of mean-preserving spreads gradually spreads out the distribution of information from the first limit case to the second. We parameterize the distribution of information by a homogeneity parameter $\kappa \in \mathbb{R}^+$ such that the distribution becomes less dispersed as κ increases, spanning the whole feasible range. That is, an increase in κ entails a mean-preserving contraction of θ_l. The limit case of maximum information heterogeneity corresponds to $\kappa = 0$ and the limit case of perfectly homogeneous information to $\kappa \to \infty$.[18]

In a decentralized system, each region forms a separate constituency with a share of informed voters θ_l. It has an independent local government that allocates the regional budget b. Local politicians with skills $\eta^D_{l,p,t}$ invest in the provision of local public goods $x^D_{l,p,t}$ and extract rent $r^D_{l,t} = b - \sum_{p=1}^{P} x^D_{l,p,t}$.

Under centralization, instead, the central government is elected by a single unified constituency whose share of informed voters equals the average across regions $\frac{1}{L}\sum_{l=1}^{L} \theta_l$. We rule out economies of scale: the central government budget equals the sum bL of the regional budgets. Central politicians with skills $\eta^C_{p,t}$ choose expenditures $x^C_{l,p,t}$ for each public good p in each region l and extract rent $r^C_t = bL - \sum_{l=1}^{L}\sum_{p=1}^{P} x^C_{l,p,t}$. The central government may be required to provide public goods uniformly across regions ($g^C_{l,p,t} = g^C_{p,t}$ for all l), either by a technological or by a constitutional constraint (Oates 1972; Alesina and Spolaore 2003). Conversely, it may be able to allocate spending across regions with complete discretion (Lockwood 2002; Besley and Coate 2003).

Different government structures admit the following ranking in terms of aggregate social welfare.

PROPOSITION 1. Aggregate social welfare is higher under decentralization than under centralization without a uniformity constraint. It is highest under centralization with a uniformity constraint if and only if preferences are sufficiently

18. For example, $\theta_l \sim B(\overline{\theta}\kappa, (1-\overline{\theta})\kappa)$ could have a beta distribution with mean $\overline{\theta}$ and sample size (i.e., confidence) κ. Our results do not rely on this particular specification.

homogeneous ($\upsilon \geq \overline{\upsilon}$). Centralization is more likely to be optimal when information is more heterogeneous ($\overline{\upsilon}$ is increasing in κ) and politicians' ability less variable ($\overline{\upsilon}$ is increasing in σ).

Centralization unambiguously reduces total rents when different regions have different information. Merging heterogeneous regions creates a single polity whose level of voter information equals the average across regions. Aggregate rent extraction declines because it is a convex function of voter information, as established in Lemma 1:

$$\frac{1}{L}\sum_{l=1}^{L}\rho(\theta_l) \geq \rho\left(\frac{1}{L}\sum_{l=1}^{L}\theta_l\right). \tag{12}$$

Nonetheless, centralization reduces welfare if the central government can operate without a uniformity constraint that requires public goods to be provided identically in all regions. Office-seeking politicians target government spending to the most politically influential regions. In our model, influence stems from information. Absent a uniformity constraint, central government spending in different regions is proportional to voter information:

$$\frac{\sum_{p=1}^{P} x_{l,p,t}^{C}}{\sum_{p=1}^{P} x_{m,p,t}^{C}} = \frac{\theta_l}{\theta_m} \text{ for all } l \text{ and } m. \tag{13}$$

This equilibrium allocation features harmful regressive redistribution. Independent local governments extract larger rents and provide fewer public goods in less informed regions. Centralization without uniformity further reduces public-good provision in these regions, increasing it instead in better informed ones. Then, aggregate social welfare declines even though the total provision of public goods rises as aggregate rent extraction falls.

On the contrary, with a uniformity constraint the decrease in rents is accompanied by progressive redistribution that raises welfare further. Centralization slightly increases rent extraction in better informed regions, but greatly reduces it in less informed ones, which have a higher marginal utility of public goods because their local government is worse. Intuitively, the uninformed gain from outsourcing government monitoring to better-informed voters in other regions. The informed can also

share in the accountability gains from centralization if a uniformity constraint is imposed on some goods but not others. Then the uninformed enjoy greater accountability in the provision of uniform public goods, and the informed enjoy greater influence over the provision of discretionary ones. Online Appendix B shows formally how such partial uniformity can make centralization a Pareto improvement, albeit at the cost of sacrificing welfare maximization.[19]

The key result in Proposition 1 is that the welfare-maximizing government structure for heterogeneous regions reflects a trade-off between greater preference matching under decentralization and greater accountability under centralization. On the one hand, the central government must be required to provide public goods uniformly, so centralization sacrifices the ability to tailor local public goods to local preferences. The more regions differ in their ideal allocation, the greater the costs of political integration. Thus, preference heterogeneity is a centrifugal force. On the other hand, rent extraction falls when the most informed regions hold politicians accountable for everyone. The more regions differ in their monitoring ability, the greater the benefits of political integration. Thus, information heterogeneity is a centripetal force.

If tastes are similar enough across regions, centralization maximizes welfare despite the absence of externalities or economies of scale. Centralization is more likely to be optimal the more information varies across regions. So long as information is not perfectly homogeneous, it is optimal when preferences are similar but not identical ($\overline{v}$ is finite).

The final result in Proposition 1 reflects the cost of uniformity in government competence. Under decentralization, each region selects—to the best of its imperfect screening ability—ruling politicians who are most talented at providing those public goods the region finds most important. The central government, instead, has average skills that try to satisfy all regions but truly fit none. When the variance of politicians' ability is greater, so is the cost of such uniformity. Then, centralization becomes less

19. Public-good spillovers across regions are another force that can make centralization a Pareto improvement. Online Appendix B shows that in our model the screening of politicians is better at the central than the local level if there are externalities. Furthermore, we provide a political agency microfoundation for the classic assumption that decentralization distorts the budget allocation for spillover-generating public goods (Oates 1972).

appealing because it distorts the allocation of talent but has no impact on average screening:

$$\sum_{p=1}^{P} \mathbb{E}\hat{\eta}_p^C = \phi\sigma^2\overline{\theta} = \frac{1}{L}\sum_{l=1}^{L}\sum_{p=1}^{P} \mathbb{E}\hat{\eta}_{l,p}^D. \tag{14}$$

This invariance, however, follows from the assumption that voter information about public goods is independent of the level of government that provides them. This assumption is realistic to the extent that voter knowledge reflects individual characteristics such as human capital, social capital, or civic engagement. Yet voter information also reflects differences in media coverage, which plausibly varies with political integration. In particular, the media are more likely to report on centralized policies because they concern a broader audience (Gentzkow 2006; Snyder and Strömberg 2010). Such an increase in reporting would entail additional efficiency gains from centralization, through better selection as well as better incentives (Glaeser and Ponzetto 2014). Then greater variance in politicians' ability might make political integration more appealing, rather than less.

Do the theoretical results in Proposition 1 have counterparts in the real world? We certainly cannot prove empirically whether the European Union or an increasing federal share of U.S. government spending is good or bad. There is, however, evidence supporting the key points in our model: discretionary spending by the central government can short-change less informed groups, decentralized control has often been associated with corruption and limited political accountability, the benefits of centralization are often greater for less informed populations, and decentralization has been more successful where accountability varies less across regions.

Strömberg (2004) studies the allocation of discretionary government spending during the New Deal and documents that state governors favored counties with a greater share of radio listeners and thus better informed voters. If we accept his identifying assumption that ground connectivity and woodland cover have no direct effect on the effectiveness of government expenditure, it follows that voter information alone is driving these differences in public spending across space. The tendency of discretionary spending to follow knowledge is precisely why Proposition 1 finds that discretion is bad.

The downsides of discretion may also explain why uniformity is common in many government policies. It may seem counterintuitive that U.S. federal housing policy should offer similar subsidies to building in areas where supply is constrained, like New York City, and areas where supply seems almost unlimited, like Houston. One explanation for spatial uniformity is that the tendency of locational discretion to harm particular regions is well understood.

The fundamental downside of decentralization in our model is that it leads to less accountability and more corruption. We know of no studies that clearly illustrate the relative corruption of national versus local governments in the United States and Europe. However, the history of U.S. state and city governments is consistent with our theoretical prediction.

At the turn of the twentieth century, the governments of large U.S. cities were infamous for their corruption. New York's "Boss" Tweed and his formidable Tammany Hall machine live on in popular memory as epitomes of organized graft in local government.[20] Other cities had equally corrupt administrations—a major theme of the progressive movement (Steffens 1904).[21] This urban experience was very far from Tiebout's (1956) and Oates's (1972) vision of local governments responding tightly to the desires of their residents.

Federal intervention eradicated the corrupt manipulation that had characterized U.S. local politics, at least in the context of welfare spending. Until the Great Depression, poverty relief managed by states and localities was a byword for patronage and graft. The New Deal—the most dramatic episode of centralization in the history of the United States—introduced strict federal oversight of welfare programs. One consequence was a striking decrease in corruption (Wallis, Fishback, and Kantor 2006).

20. The New York County Courthouse, better known as the Tweed Courthouse, became a veritable monument to corruption. Its construction took over 20 years and cost $12 million, with overbilling of comical proportions. A Tammany Hall ring member was paid $133,187 (around $2 million in present-day terms) for two days' work as a plasterer.

21. Chicago's street railways are another infamous case. The city council granted exclusive franchises on such favorable terms that in 1893 the entire system returned a mere $50,000 to the city. Instead, traction magnate Charles Yerkes spent $1 million in bribes to get through the state legislature a law enabling Chicago aldermen to grant franchises for no less than 50 years and without any compensation to the city.

While city politics cleaned up after the New Deal, state governments remained notorious for corruption (Wilson 1966). Since World War II, 10 governors and 9 members of state executives have been convicted for official corruption and sentenced to jail. No member of the federal Cabinet, let alone a president, has been charged with crimes investigated as part of the federal prosecution of public corruption.

Contemporary cross-country studies have yielded conflicting and inconclusive results on the relationship between decentralization and corruption (Treisman 2007). Historical evidence from around the world, however, shows that political integration often had a positive impact on government accountability. Centralized political institutions in precolonial Africa reduced corruption and fostered the rule of law. They caused a long-lasting increase in the provision of public goods that endured into the postcolonial period (Gennaioli and Rainer 2007). Fiscal centralization was a key element in the modernization of European states. It proved a necessary step for the consolidation of state capacity, which was in turn a critical determinant of economic and political development (Besley and Persson 2011; Dincecco 2011). Blanchard and Shleifer (2001) argue that China grew faster than Russia in recent decades thanks to the greater strength of its central government vis-à-vis local politicians.

Proposition 1 predicts not only that centralization should reduce rent extraction but that these accountability benefits should flow mostly to the least informed regions, as long as the central government enacts a uniform policy. Empirical evidence on reforms to public education systems bears out this prediction. In the early 1990s, Argentina transferred control of federal secondary schools to provincial governments. Student test scores rose in richer municipalities, but failed to rise or even fell in poor ones (Galiani, Gertler, and Schargrodsky 2008). Decentralization increased inequality and harmed those already disadvantaged. A 1998 university reform in Italy transferred responsibility for faculty hiring from the national ministry to individual universities. Faculty hires became significantly more nepotistic in provinces with low newspaper readership. Those with higher readership experienced at best a marginal improvement (Durante, Labartino, and Perotti 2014). Decentralization worsened the quality of academic recruitment and hurt the least informed regions the most.

Environmental policy in the United States also provides suggestive support for our theoretical prediction. The Clean Air Act

of 1970 transferred responsibility for pollution regulation from the state and local governments to the federal Environmental Protection Agency. Relative to preexisting trends, pollutant emissions began to decline considerably faster in states with lower newspaper circulation (we provide a formal difference-in-differences analysis in Boffa, Piolatto, and Ponzetto 2014).

The conclusion of Proposition 1 is that decentralization is desirable only if accountability is relatively homogeneous across regions. Our finding is consistent with historical evidence on the formation of unified nation-states in Germany and Italy. Both countries were unified in the second half of the nineteenth century: the Kingdom of Italy was established in 1861 and the German Empire in 1871. Before unification, Germany comprised many modern and well-functioning states. In Italy, the quality of preunitary institutions was lower and more heterogeneous. The Kingdom of Sardinia, which led the process of unification, could be considered the only efficient modern state. Consistent with our theory, these different patterns of institutional quality before unification can explain why Germany was conceived as a federal nation-state and Italy as a unitary one (Ziblatt 2006). Remarkably, both the degree of centralization and the underlying differences in accountability have remained larger in Italy than in Germany up to the present day—excepting the tragic parenthesis of German centralization under Nazism.

IV. How Many Levels of Government Should There Be?

The classic theory of fiscal federalism studies "which functions and instruments are best centralized and which are best placed in the sphere of decentralized levels of government" (Oates 1999, p. 1120). This standard approach suggests that there should be as many levels of government as there are geographic units a function is optimally tied to. Evidence from local governments in the United States, however, paints a different picture. Special-purpose districts managing individual public services for different and overlapping areas have performed poorly in terms of efficiency and accountability (Berry 2009). In this section, we explain why the proliferation of government tiers can harm welfare, and we study when it is optimal to create a federal structure in which some policy decisions are centralized and others decentralized.

We assume the same distribution of voter information as in Proposition 1, with mean $\overline{\theta}$ and a homogeneity parameter κ. However, we now consider two kinds of public goods at the opposite extremes of preference heterogeneity. First, there is a set of public goods for which all regions have perfectly homogeneous preferences ($\upsilon \to \infty$). By Proposition 1, these public goods would best be provided by a central government if there were no other policy choices. For the second set of public goods, preferences are completely idiosyncratic ($\upsilon = 0$ and $P \to \infty$). Each region benefits exclusively from its own ideal variety, and derives no utility at all from any of the $L - 1$ ideal varieties of the other regions. Absent other policies, Proposition 1 established that these idiosyncratic public goods should be provided by decentralized local governments. With both types of public goods, a resident i of region l has utility

$$u_t^i = \tilde{u}_t^i + \alpha_0 \log g_{l,0,t} + (1 - \alpha_0) \log g_{l,l,t}, \tag{15}$$

where g_0 is a composite bundle of all the homogeneously desired public goods, and g_l is region l's desired variety of idiosyncratic public goods. The ideal share $\alpha_0 \in (0, 1)$ provides a measure of preference homogeneity in this setting.

The structure of government is described by an allocation of powers and budgets to the two levels of government, local and central. As before, full decentralization means that each local government provides the residents of its region l with both the homogeneously desired public goods ($g_{l,0}$) and their ideal variety of idiosyncratic public goods ($g_{l,l}$). Conversely, the government is fully centralized if the central government is tasked with providing all public goods to residents of all regions.

An intermediate possibility is the creation of a federal system. The central government provides homogeneously desired public goods ($g_{l,0}$) to all regions, while every region has its own local government provide the idiosyncratic public good $g_{l,l}$.[22] The overall budget remains exogenously fixed at Lb. Consistent with our focus on expenditures, we assume that all regions must contribute equally to the central government budget. Its size b_C then suffices to characterize the budget allocation. Local government budgets are determined residually as $b_D = b - \frac{b_C}{L}$ for every region.

22. A federal system with the opposite allocation of powers is theoretically possible but intuitively undesirable. We prove in Online Appendix C that it can never be welfare-maximizing.

The central government may be required to provide any public good uniformly. The uniformity constraint is imposed independently on each good. It may apply to some goods and not others. It may not, however, apply to an aggregate of goods. This restriction is immediate for a technological constraint because every good is distinct. The aggregate amount of public goods provided to a region ($\sum_{p=0}^{L} g_{l,p,t}$) cannot be constrained constitutionally either. The quantities of different goods cannot be properly compared by an impartial auditor, so it is unfeasible to require the provision of "separate but equal" public goods to different regions.

The welfare-maximizing structure of government admits the following characterization.

PROPOSITION 2. A federal system is optimal if differences in voter information are large enough ($\kappa < \overline{\kappa}$) while differences in preferences are neither too small nor too large ($\alpha_0 \in (\overline{\alpha}_{D\sim F}, \overline{\alpha}_{F\sim C})$). A federal system is more likely to be optimal when information is more heterogeneous ($\overline{\alpha}_{D\sim F}$ is increasing and $\overline{\alpha}_{F\sim C}$ decreasing in κ) and politicians' ability more variable ($\overline{\kappa}$ is increasing in σ and $\frac{\partial \overline{\alpha}_{F\sim C}}{\partial \sigma} > \frac{\partial \overline{\alpha}_{D\sim F}}{\partial \sigma} = 0$).

Full centralization is optimal if differences in preferences are small ($\kappa < \overline{\kappa}$ and $\alpha_0 \geq \overline{\alpha}_{F\sim C}$, or $\kappa \geq \overline{\kappa}$ and $\alpha_0 \geq \overline{\alpha}_{D\sim C}$). Full decentralization is optimal if differences in preferences are large ($\kappa < \overline{\kappa}$ and $\alpha_0 \leq \overline{\alpha}_{D\sim F}$, or $\kappa \geq \overline{\kappa}$ and $\alpha_0 < \overline{\alpha}_{D\sim C}$). Full centralization is less likely to be optimal when politicians' ability is more variable ($\frac{\partial \overline{\alpha}_{D\sim C}}{\partial \sigma} > 0$).

Our model of accountability reverses the standard logic of fiscal federalism. The existence of some policy instruments that are best centralized and some others that are best decentralized does not immediately imply that the government should be structured on federal lines. On the contrary, if regional differences in voter information are negligible, it is optimal to have a single level of government: either only a central government, or only independent local governments. This key result reflects endogenous economies of scope in government accountability.

Politicians with little power have low-powered incentives. They control a smaller budget, so they have a lower value of holding office. Moreover, their skills have a lower impact on voters' utility, so other factors are more likely to determine their reelection. As a result, their career concerns are weaker. In

equilibrium, incumbents have incentives to demonstrate each skill in proportion to its welfare value. For example, a politician tasked with providing g_0 to voters with average information $\overline{\theta}$ invests $x_0 = \alpha_0 \overline{\theta} \phi R$ if he values reelection R. Crucially, the equilibrium value of reelection is proportional to the budget a politician controls. Then there are no economies of scale across regions: halving both the budget and the population served leads to invariant spending per capita. Instead, there are economies of scope across goods: halving both the budget and the set of public goods provided leads to lower spending on each good and a higher share of the budget dissipated as rents.

Centralization minimizes aggregate rent extraction because it exploits both these economies of scope and the efficiency benefits of delegating government monitoring to the best monitors. As in Proposition 1, however, the central government fails to match idiosyncratic local needs. Under full centralization, each region unavoidably receives its ideal variety of idiosyncratic public goods in proportion to its residents' information:

$$\frac{x^C_{l,l,t}}{x^C_{m,m,t}} = \frac{\theta_l}{\theta_m} \text{ for all } l \text{ and } m. \tag{16}$$

The optimal provision of homogeneously desired public goods is uniform across regions, so a uniformity constraint suffices to ensure it. On the contrary, requiring uniform provision of idiosyncratic public goods only makes misallocation worse. The central government keeps catering disproportionately to the preferences of the informed, but it has to provide their ideal variety to other regions that derive no benefit from it. This uniformity constraint is so wasteful it makes every region worse off than discretionary central provision of idiosyncratic public goods.

Preference heterogeneity then has a natural effect on the optimal structure of government. If preferences are highly idiosyncratic, decentralization is optimal because local governments are best at preference-matching. If preferences are highly homogeneous, centralization is optimal because only rent-minimization matters. In both extreme cases, one class of public goods is marginal, so it is worth sacrificing its optimal provision to exploit economies of scope and raise accountability in the provision of the dominant public goods.

When preference heterogeneity is intermediate, both idiosyncratic and homogeneously desired public goods are important.

The key result in Proposition 2 is that a federal system is then optimal if and only if differences in voter information across regions are large enough. When the information gap is larger, uninformed regions gain more from delegating monitoring to informed ones. Hence, there are greater benefits from having a central government provide homogeneously desired public goods ($\overline{\alpha}_{D\sim F}$ is increasing in κ). Greater heterogeneity also implies that uninformed regions lose more from ceding power to informed ones. Thus, there are greater costs of having the central government provide idiosyncratic public goods, too ($\overline{\alpha}_{F\sim C}$ is decreasing in κ).

When differences in voter information are large, it is worth sacrificing economies of scope to reap the large benefits of a progressive transfer of accountability without paying the large costs of a regressive transfer of power. Figure III represents graphically the optimal structure of government. The larger the difference in information, the larger the region F in which a federal system is optimal.

As in Proposition 1, a downside of centralization is the uniformity of central politicians' skills. Thus, greater variation in the pool of political talent reduces the appeal of full centralization. As a consequence, not only decentralization but also a federal system become more attractive.[23]

Proposition 2 established that multiple levels of government come at the cost of reduced government efficiency and accountability, even if they may be desirable for preference-matching and distributive reasons. The experience of local government in the United States empirically bears out our prediction. Many states have overlapping layers of county governments, municipal governments, and multiple special-purpose governments, such as elected school districts and independent districts managing specific public utilities. The performance record of special-purpose governments has been disappointing, and they have proved prone to capture by special interests (Berry 2009). The employees of the special-purpose district are often the key voting bloc in its elections. Public libraries provide a telling example of systematic inefficiency. When they are run by directly elected special-purpose library districts, they have larger budgets, but neither more visitors nor higher circulation. On the contrary, they hold fewer books and fewer of their employees are actually librarians.

23. In Figure III, the continuous locus $\alpha_0 = \max\{\overline{\alpha}_{F\sim C}, \overline{\alpha}_{D\sim C}\}$ shifts up, and so does its intersection $\overline{\kappa}$ with the locus $\alpha_0 = \overline{\alpha}_{D\sim F}$.

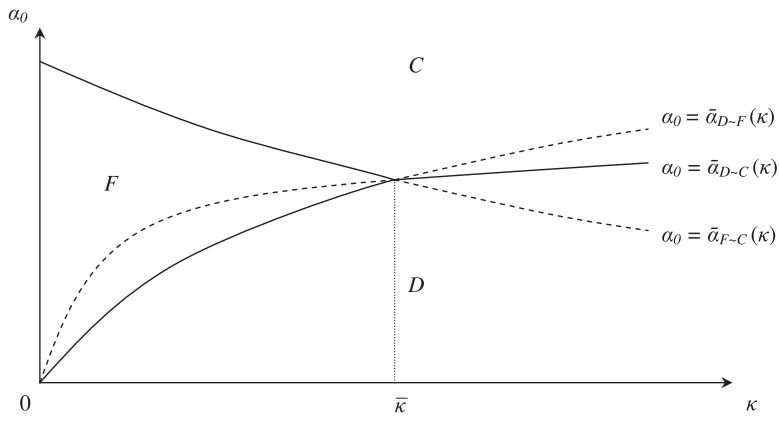

FIGURE III
Optimal Federalism

Evidence from Europe confirms that multiplying government tiers has detrimental effects. In England, local government most commonly has two levels: counties and districts. A sizable minority of areas are governed instead by a unitary authority entrusted with all local government tasks. Unitary authorities are more efficient, particularly because the two-tier structure is linked to lower labor productivity and excess employment (Andrews and Boyne 2009).

France has three nested tiers of subnational governments (regions, departments, and municipalities) plus various associations of municipalities. This complex and multilayered structure has been a source of inefficiency and institutional weakness, especially at intermediate levels (Le Galès and John 1997). In its two latest reports on local government finances, the French Court of Auditors stresses that the proliferation of subnational government tiers determines unproductive public employment. It also highlights inadequate governance mechanisms and advocates intervention by the national parliament to directly set goals and standards for local governments. Pruning the local government structure is on the French government's agenda. The Attali Commission recommended abolishing the departmental tier within 10 years. President François Hollande has proposed abolishing elected departmental councils by 2020.

In Germany, since 2000 three states (Rhineland-Palatinate, Saxony-Anhalt, and Lower Saxony) have abolished one level of

local government. Italy abolished elected provincial councils in 2014, and the government has proposed a constitutional reform to abolish provinces altogether. Italy's three-tier subnational structure (regions, provinces, and municipalities) is widely recognized as inefficient: it was arguably designed specifically as a way for political parties to provide patronage and sinecures (Dente 1988).

Cross-country evidence also supports the predictions of Proposition 2. In countries with more levels of government, firms report having to pay more frequent and costlier bribes. This positive correlation between corruption and the number of government tiers is particularly robust, and its magnitude is a first-order concern for developing countries. Fan, Lin, and Treisman (2009, p. 32) conclude that "other things equal, in a country with six tiers of government (such as Uganda) the probability that firms reported 'never' being expected to pay bribes was .32 lower than the same probability in a country with two tiers (such as Slovenia)."

Although there is clear evidence that the multiplication of government tiers dilutes accountability, we know of no equally clear evidence on the distributive benefits of federalism. Nonetheless, the pattern of political discourse in the United States is suggestively consistent with our theoretical prediction that the least informed regions benefit the most from a federal structure relative to either unitary alternative. On average, Southern states have less educated voters and lower newspaper readership. They also have more corrupt governments (Glaeser and Saks 2006). The distributive predictions of our model can then help explain why the South is at the same time particularly patriotic—for example, it provides a disproportionate share of U.S. military personnel—but also keenest on curbing the expansion of federal power and preserving the states' independent policy-making responsibilities.

When neither full centralization nor full decentralization is optimal, we can characterize the precise structure of the optimal federal system.

COROLLARY 1. In the optimal federal system, the budget, productivity and accountability of the central government are lower when differences in preferences are larger ($\frac{\partial b_C^*}{\partial \alpha_0} > 0$, $\frac{\partial \mathbb{E}\hat{\eta}_0^C}{\partial \alpha_0} > 0$ and $\frac{\partial \rho^C}{\partial \alpha_0} < 0$).

The budget, productivity, and accountability of local governments are higher when differences in preferences are larger ($\frac{\partial b_D^*}{\partial \alpha_0} < 0$, $\frac{\partial \mathbb{E}\hat{\eta}_{l,l}^D}{\partial \alpha_0} < 0$ and $\frac{\partial \rho_l^D}{\partial \alpha_0} > 0$). Rent extraction by local governments increases with differences in information ($\frac{1}{L}\sum_{l=1}^{L} \rho_l^D$ is decreasing in κ).

Overall rent extraction (by both levels of government) increases with differences in information. It is a concave function of preference heterogeneity and it reaches a maximum at the value $\breve{\alpha}_0 \in \left(0, \frac{1}{2}\right)$ for which local governments have on average the same accountability as the central government ($\alpha_0 = \breve{\alpha}_0 \Leftrightarrow \rho^C = \frac{1}{L}\sum_{l=1}^{L} \rho_l^D$). The difference in preferences associated with maximum rents increases with differences in information ($\breve{\alpha}_0$ is increasing in κ).

The comparative statics on each level of government highlight the fundamental strength of a federal system. Resources flow to the level of government where they are most useful. All regions prefer the unique efficient budget allocation that gives each level of government resources proportional to the ideal share of the public good it is responsible for providing:

$$b_C^* = \alpha_0 bL \text{ and } b_D^* = (1-\alpha_0)b. \tag{17}$$

Voter monitoring of politicians obeys a similar equilibrium allocation. Screening for competence is proportional to the welfare weight of the public goods each politician is in charge of providing:

$$\mathbb{E}\hat{\eta}_0^C = \alpha_0 \phi \sigma^2 \overline{\theta} \text{ and } \mathbb{E}\hat{\eta}_{l,l}^D = (1-\alpha_0)\phi\sigma^2\theta_l. \tag{18}$$

Hence, incentives improve and rent extraction declines when a politician has more important responsibilities:

$$\rho^C = \left(1 + \alpha_0 \frac{2\delta}{2-\delta}\phi\overline{\theta}\right)^{-1} \text{and } \rho_l^D = \left[1 + (1-\alpha_0)\frac{2\delta}{2-\delta}\phi\theta_l\right]^{-1}, \tag{19}$$

such that $\frac{\partial \rho^C}{\partial \alpha_0} < 0 < \frac{\partial \rho_l^D}{\partial \alpha_0}$.

Aggregate rent extraction is lowest when one level of government accounts for most public-good provision, so it controls most of the budget and is also the main focus of voter monitoring. Then total rents are low because one level of government is large and accountable, while the other is relatively unaccountable but

small. By Proposition 2, when this logic (and the value of α_0) is brought to an extreme, a federal structure becomes undesirable: the small and unaccountable level of government is best abolished. Hence, Proposition 1 highlights the second-best nature of the optimal government structure. Federalism is welfare-maximizing for intermediate values of α_0, but total rents are then larger, too.

Intuitively, rent extraction is highest when both levels of government are equally accountable ($\rho^C = \mathbb{E}\rho_l^D$). Then, if either grew more important it would control a larger budget share and extract proportionally fewer rents from it. Rents are largest when the central government is smaller than the local ones ($\check{\alpha}_0 < \frac{1}{2}$). This is a natural consequence of greater accountability at the central level in the presence of heterogeneous information. As differences in voter information grow larger, so does the inefficiency of local governments, and thus of a federal system that includes them. Accordingly, the peak of rent extraction is associated with a greater importance of local governments.

V. What Should Determine the Boundaries of Governments?

Government structure is not entirely described by the number of tiers. The size of subnational jurisdictions can also vary. Is it better to have few large local governments or many small ones? Our model can be applied directly to study the optimal boundaries of governments. Proposition 1 considered a simple symmetric setting in which either all regions should integrate or each should have its independent government. The intuition generalizes to asymmetric cases. Regional boundaries should be drawn so that people with similar preferences but different information share a government, whereas those with different preferences but similar information do not.

In this section we extend our model by relaxing the assumption that voters are sorted into geographic regions with internally homogeneous preferences. To study optimal boundaries when ideological groups do not naturally coincide with geographic regions, we assume a simple twofold partition of voters by ideology and information.

Voters have ideological preferences for two distinct public goods L and R. Left-wingers desire the former and have utility

$u^i_{L,t} = \tilde{u}^i_t + \log g_{l,L,t}$. Right-wingers desire the latter and have utility $u^i_{R,t} = \tilde{u}^i_t + \log g_{l,R,t}$. This simple preference structure provides a stylized model of local government consistent with Proposition 2. Preferences over locally provided public goods are highly heterogeneous because public goods that all voters desire homogeneously should be provided by the federal government instead.

Each ideological group comprises voters with different levels of information. Better informed voters succeed at inferring the incumbent's competence from realized policy outcomes with probability θ_I. Relatively uninformed voters have a lower probability of learning $\theta_U < \theta_I$.

A country is then characterized by the sizes of the four groups $\lambda_{L,I}$, $\lambda_{L,U}$, $\lambda_{R,I}$, and $\lambda_{R,U}$. We consider partitions of this overall population into autonomous regions or federal states. Each region is endowed with a budget of b units per resident, so there are no economies of scale. Moreover, a region is the minimal administrative unit, so the regional government is subject to a technological uniformity constraint: it cannot differentiate the provision of public goods across residents.

We begin by characterizing the optimal regional structure when there are no constraints on how citizens can be partitioned into regions.

PROPOSITION 3. Optimal regions are perfectly separated by preferences and perfectly mixed by information (every region l has either $\lambda_{l,L,I} = \lambda_{l,L,U} = 0$ and $\frac{\lambda_{l,R,I}}{\lambda_{l,R,U}} = \frac{\lambda_{R,I}}{\lambda_{R,U}}$, or $\lambda_{l,R,I} = \lambda_{l,R,U} = 0$ and $\frac{\lambda_{l,L,I}}{\lambda_{l,L,U}} = \frac{\lambda_{L,I}}{\lambda_{L,U}}$).

Without exogenous constraints, the optimal partition intuitively resolves the two forces highlighted by Proposition 1. Preference heterogeneity is a centrifugal force that can be accommodated by separating groups with different ideal allocations. Such optimal segregation reflects Tiebout's (1956) classic intuition. It is typically optimal when there are no economies of scale and no constraints on creating as many regions as there are desired bundles of public goods (Bewley 1981). The novelty of our model lies in the centripetal force caused by differences in information. A partition that achieves homogeneous preferences within each region can nonetheless be suboptimal. Optimality also requires the perfect mixing of like-minded voters with different levels of information. Citizens suffer from sharing a government with others with opposite preferences who cause a

distributional conflict. They suffer no less from being cut off from better-informed voters with the same preferences, whose influence is necessary to keep the local government accountable.

Proposition 3 highlights that an ideologically homogeneous but uniformly ill-informed region is plagued by bad governance. Its government reflects the preferences of local residents, but it is also unaccountable, inefficient, and corrupt. This prediction of our model is consistent with evidence from local governments in the United States. City politicians have at times succeeded in creating large local majorities of their poorer and less educated supporters by encouraging the out-migration of a rival higher-status group. The detrimental consequences of his process are best illustrated by the long career of Boston mayor James Michael Curley (Glaeser and Shleifer 2005). Both his policies and his stark rhetoric championed the poor Irish community against the richer Anglo-Saxon Protestants who had previously dominated the city. The end of Brahmin dominance pleased Boston's Irish and removed the discrimination they had suffered from. However, Curley's administration was inefficient and corrupt; Boston declined under his government. Similar patterns emerge in other cases of populist local politics catering to particular ethnic and socioeconomic constituencies, such as African Americans in Detroit under Coleman Young.

The optimal partition described by Proposition 3 has two contrasting features. Tension between the two can entail a welfare loss when groups with different preferences are separated. Proposition 1 characterized one set of circumstances leading to this outcome. When voters' preferences are not completely distinct, separation is undesirable if differences in voter information are large enough. Another possibility is that perfect separation à la Tiebout is technologically impossible because residents with different preferences are mixed in a narrow area such as a city or county. If perfect separation is impossible, is partial separation desirable, or is it even worse than perfect integration?

Consider two symmetric atomistic locations. Their total population is identical, but the first location has a majority of left-wing residents and the second a majority of right-wing residents. The distribution of the population is characterized by a degree of ideological sorting $\tau \in (0, 1)$ such that

$$\lambda_{1,L} = \lambda_{2,R} = \frac{1+\tau}{4} \text{ and } \lambda_{1,R} = \lambda_{2,L} = \frac{1-\tau}{4}. \tag{20}$$

In the limit as $\tau \to 0$ residents with different preferences are perfectly mixed, while in the limit as $\tau \to 1$ there is perfect sorting.

Voter information is also symmetric, but not homogeneous across locations. Voters with either preferences have an average probability θ of being informed in the location in which they belong to the majority. In the location where they are a minority, their information is reduced to $\theta(1 - \zeta)$ for a coefficient $\zeta \in (0, 1)$ of information disadvantage. The lower information of the minority reflects endogenous media slant (Gentzkow and Shapiro 2010). Local media choose an ideological bias to match the preferences of the local majority. As a consequence, news consumption becomes more appealing for the majority and less for the minority.

The following result characterizes formally whether political integration or partial separation is optimal when perfect segregation by preferences is impossible.

PROPOSITION 4. Aggregate social welfare is higher under political integration than under separation if minorities suffer from a high information disadvantage ($\zeta \geq \overline{\zeta}$). Integration is more likely to be optimal when ideological sorting is less complete ($\frac{\partial \overline{\zeta}}{\partial \tau} > 0$) and politicians' ability less variable ($\frac{\partial \overline{\zeta}}{\partial \sigma} > 0$).

Intraregional heterogeneity entails a new trade-off. The centripetal force is information heterogeneity of a different kind than the one underlying Proposition 1. In Proposition 4 there are no differences in average information across regions, so aggregate rent extraction is invariant. There are, however, differences in information between the majority and the minority within each location. Under separation, uninformed minorities are dominated by better informed local majorities. Political integration restores even power to the two ideological groups. Each uninformed minority gains political influence thanks to the like-minded informed majority in the other location. Thus, political integration can raise welfare even if the efficiency gains from delegated monitoring are absent.

These distributive welfare gains are monotone increasing in the information disadvantage of the minority. If information is homogeneous, separation is the constrained optimum ($\overline{\zeta} > 0$). Imperfect ideological segregation remains costly, and minorities bear a greater share of this cost. Yet political integration merely worsens overall preference matching. At the opposite extreme, if a minority is completely uninformed, it is essentially

disenfranchised. Then utilitarian welfare maximization requires political integration to protect the minority ($\overline{\zeta} < 1$ for all $\tau < 1$).

Ideological sorting provides a countervailing centrifugal force. As groups with opposite preferences are more and more segregated, the difference in preferences across regions increases and so does the appeal of political separation. In the limit, political separation is always optimal if ideological sorting is complete, as Proposition 3 already established ($\lim_{\tau \to 1} \overline{\zeta} = 1$). Finally, as in Proposition 1, greater variance in politicians' ability makes integration less attractive because of distortions in the allocation of talent.[24]

Our results speak directly to proposals for the partition of California, which have been put forward several times—most recently, venture capitalist Tim Draper attempted to introduce a ballot initiative for 2016 to split the state in six parts. The third largest state in the union by area (and largest by population), California is composed of several distinct regions. The most salient political divide is between east and west. The differences are both partisan and ideological: western California is more liberal, even among Republican voters and politicians; eastern California considerably more conservative (Kousser 2009). At a first glance, such a political divide might suggest that a break-up of coastal and inland California would be optimal on preference-matching grounds.

Proposition 4, however, cautions against this superficial assessment. Both the southeastern Inland Empire and the San Joaquin Valley contain a large Hispanic population that overwhelmingly prefers the Democratic Party. This group is much less educated, less politically knowledgeable, and less likely to vote than Republican supporters in the region, who are on average older, more likely to be white, and wealthier.[25] At the same time, the left-wing Hispanic working class in the Valley shares the political leanings of highly educated liberals on the coast. This

24. The effect of political integration on screening would be opposite if majorities were systematically less informed than minorities. Aside from comparative statics, however, the trade off presented by Proposition 4 remains in this less intuitive case. If an uninformed local majority is dominated by an informed minority, a fortiori political integration has the benefit of equalizing the power of the two groups. It raises welfare if and only if sorting is sufficiently imperfect.

25. Hispanic immigrants are also more likely not to have the right to vote, but a substantial majority of Hispanic residents of southeastern California are U.S. citizens.

ideological alignment goes beyond mere partisanship and includes shared preferences over policies: "whether they ride in limousines, Volvos, or buses, Democrats in the blue areas of the state share similar policy views" (Kousser 2009, p. 2).

As a consequence, our model suggests that the political integration of California is welfare maximizing. For relatively uneducated inland minorities to have a government corresponding to their preferences, it is essential that they share a state with ideologically aligned liberal elites in the Bay area. Right-wing Californians, instead, are sufficiently educated and influential to have a voice in state-wide politics, despite being in the minority: California had a Republican governor for 21 of the past 30 years.

The lesson of Proposition 4 applies more broadly. Disadvantaged ethnic minorities—who are less educated and often politically underrepresented—whenever possible should belong to the same polity as better-educated and higher-status voters with similar political preferences. Only then are politicians effectively held accountable to both groups.

VI. Conclusion

Is government decentralization the right answer to differences across regions? The idea has gained wide currency, from European Union law enshrining the principle of subsidiarity to independence movements in Québec, Scotland, or Catalonia and recurring proposals to split California into separate liberal and conservative states. The classic theory of fiscal federalism supports and formalizes the intuitive appeal of this notion: according to Oates's (1972) seminal decentralization theorem, decentralization is more efficient than centralization whenever regions are not identical and there are no policy spillovers.

This article offers a different perspective by focusing on a key overlooked dimension of regional heterogeneity: voters' ability to monitor politicians and hold them accountable. Our model explains why decentralization has often failed to deliver the accountability benefits anticipated by its proponents and why it is more suitable for countries with homogeneous institutional quality, like Germany, than countries with gaping regional disparities, like Italy. When voter information varies across regions, centralization yields accountability gains. The central

government is monitored mainly by the most informed regions and as a result it has better incentives than the average local government. At the same time, however, its incentives are to serve the informed and neglect the uninformed, so it must be forced to provide at least some public goods uniformly to avoid unacceptable distributive distortions. The same force thus drives both sides of a trade-off: preference heterogeneity prompts decentralization, but information heterogeneity prompts centralization instead.

As a result, the borders of governments should not reflect only the classic Tiebout (1956) logic of sorting by preferences. It is also crucial to ensure diversity of information because uninformed voters are caught between the hammer of unaccountable politicians and the anvil of better informed voters with contrasting policy priorities. The solution is for them to share a government with highly informed voters with similar tastes. Thus, California should not be broken up: the benefits of separating the liberal local majority on the coast from the conservative local majority inland seem smaller than those of grouping together the coastal liberal elite with the working-class left-wing minority in the Central Valley.

Our analysis hints that the main problem with boundaries in the United States is not that states like California are too big and diverse, but on the contrary that many states are too small. In our theory, the costs and benefits of fragmentation are driven by observables: respectively differences in voter information and in political preferences. As a first step in bringing our model to the data, we computed a rough estimate of the net benefits from merging any pair of contiguous U.S. states. We proxied the share of informed voters by that of college graduates and preferences by presidential vote shares. This simple quantitative exercise suggests that merging the smallest states in the Northeast (Delaware, Rhode Island, Vermont) and in the Mountain West (Idaho, Wyoming) with their larger neighbors would yield efficiency gains at a negligible cost in terms of preference-matching. Reuniting Virginia and West Virginia seems most attractive: the two states have very similar party vote shares, but very different levels of human capital. Our rough estimate of the welfare gains from a merger has the same order of magnitude as a permanent increase in the annual growth rate of real income per capita by 10 basis points.[26]

26. The full details of our quantitative exercise are available on request.

Our framework also offers new insights on federal systems with multiple levels of government. The standard logic of fiscal federalism suggests there should be many government layers, so that every policy instrument is tied to its optimal geographic unit. Instead, our theory shows economies of scope in government accountability. A unitary government that controls a large budget and multiple policy instruments suffers less from moral hazard than many special-purpose governments, each controlling a specific policy and separate budget. Our model thus explains why the multiplication of government tiers is empirically associated with inefficiency and poor accountability.

Furthermore, we have found that a federal structure can be desirable only if information heterogeneity is large enough. This result sounds a note of caution against the embrace of federalism as an answer to independence movements. Devolution has so far been the preferred strategy in Belgium, Spain, and the United Kingdom. However, if English and Scottish voters are equally good at monitoring government performance but prefer different government agendas, our model suggest that British federalism could be an inferior alternative either to the old model of centralization in Westminster or to full Scottish independence.

Conversely, our analysis shines a positive light on the European Union. Stark differences in institutional quality across member states are perceived as a major problem since the start of the euro crisis. How can the union include both virtuous "core" countries like Germany, the Netherlands, or Finland, and the troubled euro "periphery" of Greece, Italy, Portugal, and Spain? Our model shows that such differences in government accountability are in fact a motivating strength of the European project. They explain why we can expect efficiency gains from transferring powers to EU institutions, but also why substantial policy choices should remain at the national level.

In this article we developed a theoretical framework rather than focusing on concrete policy instruments, but the allocation of specific policies to different levels of government is clearly an important topic for future research. In this context, our theory may help explain an enduring puzzle: why the European Union does exactly what it does (Alesina, Angeloni, and Schuknecht 2005). The division of powers between member states and European institutions is not fully explained by classic considerations of externalities and preference heterogeneity. Our model shows that other considerations are equally crucial. Efficiency is

maximized by centralizing policies whose understanding by voters varies most widely across countries. Political feasibility may require striking a balance between policies that transfer power to the core and others that transfer accountability to the periphery.

FREE UNIVERSITY OF BOLZANO
INSTITUT D'ECONOMIA DE BARCELONA, UNIVERSITAT DE BARCELONA
CREI, UNIVERSITAT POMPEU FABRA, AND BARCELONA GSE

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

REFERENCES

Alesina, Alberto, Ignazio Angeloni, and Ludger Schuknecht, "What Does the European Union Do?," *Public Choice*, 123 (2005), 275–319.
Alesina, Alberto, and Enrico Spolaore, *The Size of Nations* (Cambridge, MA: MIT Press, 2003).
Alesina, Alberto, and Guido Tabellini, "Bureaucrats or Politicians? Part II: Multiple Policy Tasks," *Journal of Public Economics*, 92 (2008), 426–447.
Andrews, Rhys, and George Boyne, "Size, Structure and Administrative Overheads: An Empirical Analysis of English Local Authorities," *Urban Studies*, 46 (2009), 739–759.
Ashworth, Scott, and Ethan Bueno de Mesquita, "Electoral Selection, Strategic Challenger Entry, and the Incumbency Advantage," *Journal of Politics*, 70 (2008), 1006–1025.
Banks, Jeffrey S., and Rangarajan K. Sundaram, "Optimal Retention in Agency Problems," *Journal of Economic Theory*, 82 (1998), 293–323.
Bardhan, Pranab, and Dilip Mookherjee, "Decentralization, Corruption, and Government Accountability," in *International Handbook on the Economics of Corruption*, Susan Rose-Ackerman, ed. (Cheltenham: Edward Elgar, 2006).
Baron, David P., "Electoral Competition with Informed and Uninformed Voters," *American Political Science Review*, 88 (1994), 33–47.
Barro, Robert J., and Jong Wha Lee, "A New Data Set of Educational Attainment in the World, 1950–2010," *Journal of Development Economics*, 104 (2013), 184–198.
Berry, Christopher B., *Imperfect Union: Representation and Taxation in Multilevel Governments* (Cambridge: Cambridge University Press 2009).
Besley, Timothy, and Anne Case, "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," *American Economic Review*, 85 (1995), 25–45.
Besley, Timothy, and Stephen Coate, "Centralized Versus Decentralized Provision of Local Public Goods: A Political Economy Approach," *Journal of Public Economics*, 87 (2003), 2611–2637.
Besley, Timothy, and Torsten Persson, *Pillars of Prosperity: The Political Economics of Development Clusters* (Princeton, NJ: Princeton University Press, 2011).
Bewley, Truman F., "A Critique of Tiebout's Theory of Local Public Expenditures," *Econometrica*, 49 (1981), 713–740.

Blanchard, Olivier, and Andrei Shleifer, "Federalism With and Without Political Centralization: China versus Russia," *IMF Staff Papers*, 48 (2001), 171–179.
Boffa, Federico, Amedeo Piolatto, and Giacomo A. M. Ponzetto, "Centralization and Accountability: Theory and Evidence from the Clean Air Act," CEPR Discussion Paper No. 9514, 2014.
Dente, Bruno, "Local Government Reform and Legitimacy," in *The Dynamics of Institutional Change: Local Government Reorganization in Western Democracies*, Bruno Dente and Francesco Kjellberg, eds. (London: Sage, 1988).
Dincecco, Mark, *Political Transformations and Public Finances* (Cambridge: Cambridge University Press, 2011).
Durante, Ruben, Giovanna Labartino, and Roberto Perotti, "Academic Dynasties: Decentralization and Familism in the Italian Academia," NBER Working Paper No. 17572, 2014.
Fan, C. Simon, Chen Lin, and Daniel Treisman, "Political Decentralization and Corruption: Evidence from Around the World," *Journal of Public Economics*, 93 (2009), 14–34.
Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky, "School Decentralization: Helping the Good Get Better, but Leaving the Poor Behind," *Journal of Public Economics*, 92 (2008), 2106–2120.
Gennaioli, Nicola, and Ilia Rainer, "The Modern Impact of Precolonial Centralization in Africa," *Journal of Economic Growth*, 12 (2007), 185–234.
Gentzkow, Matthew, "Television and Voter Turnout," *Quarterly Journal of Economics*, 121 (2006), 931–972.
Gentzkow, Matthew, and Jesse M. Shapiro, "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica,* 78 (2010), 35–71.
Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, "Do Institutions Cause Growth?," *Journal of Economic Growth*, 9 (2004), 271–303.
Glaeser, Edward L., and Giacomo A. M. Ponzetto, "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions," *Journal of Public Economics*, 116 (2014), 89–105.
Glaeser, Edward L., and Raven E. Saks, "Corruption in America," *Journal of Public Economics*, 90 (2006), 1053–1072.
Glaeser, Edward L., and Andrei Shleifer, "The Curley Effect: The Economics of Shaping the Electorate," *Journal of Law, Economics, and Organization*, 21 (2005), 1–19.
Grossman, Gene M., and Elhanan Helpman, "Electoral Competition and Special-Interest Politics," *Review of Economic Studies*, 63 (1996), 265–286.
Harstad, Bård, "Harmonization and Side Payments in Political Cooperation," *American Economic Review*, 97 (2007), 871–889.
Kousser, Thad, "How Geopolitics Cleaved California's Republicans and United Its Democrats," *California Journal of Politics and Policy*, 1 (2009), 1–13.
Le Galès, Patrick, and Peter John, "Is the Grass Greener on the Other Side? What Went Wrong with French Regions, and the Implications for England," *Policy & Politics*, 25 (1997), 51–60.
Lockwood, Ben, "Distributive Politics and the Costs of Centralization," *Review of Economic Studies*, 69 (2002), 313–337.
———, "The Political Economy of Decentralization," in *Handbook of Fiscal Federalism*, Ehtisham Ahmad and Giorgio Brosio, eds. (Cheltenham: Edward Elgar, 2006).
Oates, Wallace E., *Fiscal Federalism* (New York: Harcourt Brace Jovanovich, 1972).
———, "An Essay on Fiscal Federalism," *Journal of Economic Literature*, 37 (1999), 1120–1149.
Ponzetto, Giacomo A. M., "Heterogeneous Information and Trade Policy," CEPR Discussion Paper No. 8726, 2011.
Ponzetto, Giacomo A. M., and Ugo Troiano, "Social Capital, Government Expenditures, and Growth," CEPR Discussion Paper No. 9891, 2014.
Salmond, Alex, "Preface," in *Scotland's Future: Your Guide to an Independent Scotland,* produced by APS Group Scotland (Edinburgh: Scottish Government, 2013).

Seabright, Paul, "Accountability and Decentralization in Government: An Incomplete Contracts Model," *European Economic Review*, 40 (1996), 61–89.
Snyder, James M., and David Strömberg, "Press Coverage and Political Accountability," *Journal of Political Economy*, 118 (2010), 335–408.
Steffens, Lincoln, *The Shame of the Cities* (New York: McClure, 1904).
Strömberg, David, "Radio's Impact on Public Spending," *Quarterly Journal of Economics*, 99 (2004), 189–221.
Svensson, Jakob, "Eight Questions about Corruption," *Journal of Economic Perspectives*, 19 (2005), 19–42.
Tiebout, Charles M., "A Pure Theory of Local Public Expenditures," *Journal of Political Economy*, 64 (1956), 416–424.
Treisman, Daniel, *The Architecture of Government: Rethinking Political Decentralization* (Cambridge: Cambridge University Press, 2007).
Wallis, John J., Price Fishback, and Shawn Kantor, "Politics, Relief, and Reform: Roosevelt's Efforts to Control Corruption and Political Manipulation during the New Deal," in *Corruption and Reform: Lessons from America's Economic History*, Edward L. Glaeser and Claudia Goldin, eds. (Cambridge, MA: National Bureau of Economic Research, 2006).
Wilson, James Q., "Corruption: The Shame of the States," *Public Interest*, 2 (1966), 28–38.
Ziblatt, Daniel, *Structuring the State: The Formation of Italy and Germany and the Puzzle of Federalism* (Princeton, NJ: Princeton University Press, 2006).