

Political decentralization and policy experimentation

Hongbin Cai^a

Daniel Treisman^b

Since 1932, when Justice Louis Brandeis remarked that in a federal system states can serve as “laboratories” of democracy, political decentralization has been thought to stimulate policy experimentation. We reexamine the political economy behind this belief, using a simple model of voting in centralized and decentralized democracies. We find the electoral logic suggests the opposite conclusion: centralization usually leads to “too much” policy experimentation, compared to the social optimum, while decentralization leads to “too little”. Three effects of centralization—an “informational externality”, a “risk-seeking” effect, and a “risk-conserving” effect—account for the different outcomes.

January 2009

^a Guanghua School of Management, Peking University, Beijing, China, 100871.

^b Corresponding author. Department of Political Science, University of California, Los Angeles, 4289 Bunche Hall, Los Angeles CA 90095-1472, treisman@polisci.ucla.edu,
Tel: 310 968 3274, Fax: 310 825 0778

H11, H70, H77, H83, D72, D78

Keywords: decentralization, innovation, policy experiments, voting, information.

We thank Pablo Beramendi, Tim Besley, David Canon, Scott Desposato, Charles Franklin, Scott Gehlbach, John Gerring, Clark Gibson, Martin Gilman, Vladimir Gimpelson, Roger Gordon, Torben Iversen, Herbert Kitschelt, Thad Kousser, Anirudh Krishna, Matt McCubbins, Dilip Mookherjee, Karen Remmer, Jonathan Rodden, Phil Roeder, Matt Singer, Aseema Sinha, David Soskice, David Weimer, the editors and anonymous reviewers of the QJPS, and participants in seminars at Duke University, UCSD, Columbia University, University of Wisconsin, Wissenschaftszentrum Berlin, and the Higher School of Economics, Moscow.

Since Justice Brandeis' famous remark in 1932 that in a federal system states can serve as laboratories to test novel policies, political decentralization has been widely thought to stimulate policy experimentation and innovation.¹ The notion has been influential in American jurisprudence (Greve 2001). In the Supreme Court, Brandeis' claim has been used to defend letting states set policy on everything from physician-assisted suicide and medical marijuana to jury trial procedure and gun-free school zones.² The argument is also common in economics and political science. In his classic work on fiscal federalism, Wallace Oates suggested that one of the three main benefits of decentralization is that it may "result in greater experimentation and innovation in the production of public goods."³

This paper reexamines the political economy behind this argument. Why might decentralization increase the frequency of policy experimentation? Some assume that in a centralized system the government simply cannot differentiate its policies geographically (e.g. Strumpf 2002). Given this, in a country with 50 regions, a centralized government can conduct only one fiftieth as many experiments per period as local governments acting autonomously. Brandeis seemed to have this in mind. Federalism, he wrote, permitted "courageous" states to experiment "without risk to the rest of the country," implying that in non-federal orders experiments must impose risks nationwide.

However, as many scholars have noted, centralized governments can implement different

¹ See Brandeis' dissent in *New State Ice Co. v. Liebmann* (1932, 285 U.S. 262). Others had expressed similar ideas before; see Bryce (1888, p.353, quoted in Oates 1999, p.1132) and Laski (1921, p.52).

² On physician-assisted suicide, see Justice O'Connor's opinion in *Washington v. Glucksberg* (521 US 702, 1997); on medical marijuana, see Justice Stevens in *US v. Oakland Cannabis Buyers' Cooperative* (532 US 483, 2001); on jury trial procedure, see Justice Powell in *Johnson v. Louisiana* (406 US 356, 1972); on gun-free schools, see Kennedy in *US v. Lopez* (514 US 549, 1995). These are discussed in Althouse (2004).

³ Oates (1972, p.12). See also his discussion in Oates (1999, pp.1131-34).

policies in different regions, and they do so all the time (Brennan and Buchanan 1980, Breton 1996, Seabright 1996, Oates 1999, Besley and Coate 2003, Lockwood 2002). Even Stalin's totalitarian regime provided native language schooling in the non-Russian Soviet republics (Bilinsky 1968). The United Kingdom and France, considered among the most politically centralized democracies, both differentiate policies geographically in numerous ways. The British operate separate legal systems for England and Scotland. In France, even in the *dirigiste* 1960s the national economic plan broke down into varied regional plans (MacLennan 1965).

Explicitly experimental local policies occur in both centralized dictatorships and centralized democracies. In the USSR, Brezhnev authorized regional economic experiments, and extended successful ones to other areas.⁴ Similar localized experiments occurred in China under both Mao and his successors. Among democracies, the UK government frequently tests policies in selected districts before rolling them out nationwide. One 2003 survey identified “well over 100” such pilot schemes conducted in the previous five years, and even worried central authorities might run out of test sites.⁵ Experiments examined financial incentives to stay in school (in 15 areas), aid to low-income workers (eight pilot areas), anti-smoking initiatives (26 zones), advisers to help single parents get jobs (eight areas), real estate market improvements (one city), and treatment sentences for minor drug offenses (four towns).⁶ Sites were chosen to ensure appropriate controls or to examine how local conditions affected outcomes.⁷

⁴ In Soviet Georgia in the 1970s, First Secretary Eduard Shevardnadze was authorized to introduce Hungarian-style reforms in agricultural management. “The experiment, which resulted in spectacular increases in agricultural production, was extended to other regions of the republic and became the model for so-called RAPOs (agricultural-industrial associations), created at the national level in 1982” (Ekedahl and Goodman 1997, chapter 1). Shevardnadze was also allowed to experiment with private commerce.

⁵ “With the growth in the number of locations that have been selected as either test or control areas for one pilot or another... the supply of suitable ‘untouched’ localities may soon be exhausted” (Jowell 2003).

⁶ See UK Office of the Deputy Prime Minister (2002), Eley et al. (2002), Jowell (2003, pp.18-30).

⁷ Our point here is to establish only that centralized governments can and often do conduct local experiments. Whether such experiments are generally more frequent under centralization—as suggested by the logic analyzed in this paper—or decentralization awaits systematic statistical investigation. A particularly compelling type of evidence would come from examining the frequency of experimentation before and after countries underwent major decentralizations.

Clearly, localized policy experiments are possible under both centralized and decentralized orders. Whether they are more frequent under one than the other needs to be explained in terms of the incentives constitutional structures create for the policymakers. We develop a model to study this question in democracies. Candidates compete for office by credibly promising to enact policies, which can be either experimental or those of the status quo. Under *decentralization*, a separate election is held in each local district, and the winner sets policy there. Under *centralization*, candidates compete for votes nationwide by proposing a set of local policies, one for each district. We derive the equilibrium number of local experiments under each system, and compare this to the level that maximizes the total expected surplus of all citizens.⁸

Contrary to the common intuition, centralization in this model usually leads to at least as many local policy experiments as decentralization—and often more.⁹ Compared to the social optimum, centralization tends to produce too much experimentation, while decentralization generally leads to too little. Several effects combine to produce this result. One, previously noted, is a positive *information externality*. Because one locality's discoveries benefit others, uncoordinated local governments tend to under-invest in experimentation. A central incumbent, who needs votes in more than one district, will at least partly internalize such externalities.

Two new effects follow from the different electoral logic for central and local candidates. Each local candidate must win a majority in his own district, but a central candidate can get elected despite losing in some units. This has two main consequences. First, since candidates for central office need only win voters from a subset of districts, they can afford to undertake very risky experiments in the others. Indeed, competition will drive them to sacrifice the units outside their coalition in search of information that might benefit those inside it. This *risk-taking effect* suggests that, when the expected value of experiments is relatively low, there will be more experimentation

⁸ For simplicity, we assume the minimum scale of a policy experiment corresponds to the local unit. All that is strictly necessary is that the national jurisdiction be divisible into more non-overlapping test sites than each local jurisdiction.

⁹ The only equilibrium in which centralization leads to less experimentation than decentralization occurs when politicians play mixed strategies and when parameter values fall in a narrow range.

under centralization than decentralization, and sometimes more than is socially optimal. By contrast, when experiments have a relatively high expected value but a high risk of failure, it may be socially optimal for all to experiment. However, voters in a candidate's support coalition may still prefer him to concentrate the risk on other units. This *risk-conserving effect* suggests that, when experiments are risky, centralization may lead to fewer experiments than is socially optimal, although—because of the other effects—usually at least as many as under decentralization.

We study several extensions. First, as the number of units increases, the temptation for local governments to free-ride on others' experimentation grows, while the risk-taking and risk-conserving effects drive a centralized government toward experimenting in half the units. As a result, when the number of units is large, centralized governments always experiment more than decentralized ones. Second, when experiments are heterogeneous or correlated, decentralization suffers from another kind of coordination failure: all units duplicate each other's efforts, choosing the same, most-promising experiment in a given round. By contrast, the central government picks a variety of experiments to increase the chance of useful discoveries. Third, if experiments impose negative externalities on other units, centralization—by internalizing these—may reduce experimentation. However, preventing such cost-exporting experiments should increase welfare.

Of course, political decentralization may have a number of benefits besides its purported impact on policy innovation. Some of these might help in experimentation as well as in other spheres. For instance, some argue that local elected officials are better informed about local conditions than are central authorities. This might enable them to select experiments better suited to their district. Others contend that local governments will suffer from fewer agency problems than a nationwide administrative hierarchy, so locally chosen experiments will be implemented more effectively. These information and agency cost arguments are, themselves, open to debate. But if correct and general, they might indeed suggest reason to let local governments experiment.

Several previous papers have modeled the relationship between decentralization and experimentation. In a pioneering contribution, Susan Rose-Ackerman (1980) showed, as we do,

that experimentation by uncoordinated local governments can result in free-riding and wasteful duplication. Our approach differs in that we explicitly model different political incentives to experiment under centralized and decentralized democracy, and contrast the equilibrium outcomes. We reproduce some effects noted by Rose-Ackerman, but also identify additional ones (e.g., the risk-taking and risk-conserving effects) and show how these interact. Strumpf (2002) identifies the informational externality. However, he assumes central governments cannot differentiate their policies geographically. We consider this unrealistic, and examine the logic if the assumption is dropped. Kollman, Miller and Page (2000) study the effectiveness of policy search in unitary and federated organizations. They model a tradeoff between assigning search to a single, sophisticated central organization and assigning it to multiple subunits that can use different strategies in parallel. Sah and Stiglitz (1986) also study the relative advantages of organizing screening agents in parallel or in a hierarchy. However, since even the most centralized state can instruct its agents to conduct multiple, parallel experiments using different algorithms, these papers are really about the center's choice of strategy rather than the comparison between centralized and decentralized constitutions.

To focus on the direct effects of state structure on innovation, we abstract from various complicating issues. For instance, we abstract from redistribution, both within and across regions. Thus, all residents of a region share the same preferences over local experimentation, and no interregional redistributive transfers are used to finance experiments.¹⁰ We also suppose governments—both local and central—are unitary actors, and, in the basic model, can commit to policies before the election.¹¹ Assuming unitary governments is an obvious simplification. In reality, many decisions at both central and local levels are made by legislatures—parliaments, town councils, etc.—whose members represent groups with different interests. In such cases, the rules and norms that govern collective choice in legislatures will clearly affect outcomes. An interesting,

¹⁰ For interesting recent analyses of redistribution in decentralized settings, see Lockwood (2002) and Bordignon, Manasse, and Tabellini (2001).

¹¹ In a longer working paper version, available at <http://www.polisci.ucla.edu/faculty/treisman>, we present an adaptation with retrospective voting and no pre-commitment, and derive similar results.

although complicated, extension that we do not consider here would be to compare the equilibrium experimentation of central and local legislatures. Some scholars have adopted a different strategy, comparing decisionmaking by a central legislature that is vulnerable to certain pathologies (universalistic spending norms, tendency to deadlock) with decisionmaking by unitary local governments (e.g. Besley and Coate 2003). To avoid stacking the deck in favor of one system or the other, we choose to assume the same decisionmaking apparatus at both levels. Assuming less efficient institutions at the center, one can of course derive advantages for decentralization, but we are interested here in how the two systems compare given similar political institutions.

1 The basic model

A state is divided into three local districts, indexed by i , each of which contains the same number of voters. Under *decentralization*, a separate local government in each district chooses policy for that unit; under *centralization*, a single central government sets the policies for each of the three districts. For a given district in a given period, the relevant decisionmaker can choose the status quo policy, A , which has consequences known to all. Or it can choose an experimental policy, $E \in \{E_1, E_2, \dots, E_l\}$, from a set of l possible experiments, where $l > 0$ is large. Any experimental policy, E_i , may have either of two outcomes—success or failure. The outcome is not known in advance, but all know the ex ante probability of success, $q_i \in (0,1)$.¹²

At the start of play, elections are held, followed by two periods in which policies are enacted.¹³ Under decentralization, each district holds a separate election to choose its local

¹² The model focuses on a single policy dimension, so only one experiment per period can occur in each district. In reality, districts may experiment on many dimensions at once. So long as these do not interact, our model applies to each of them separately. If they do interact, this will affect the frequency of experiments under centralization and decentralization in ways that do not imply an obvious advantage for either.

¹³ One might wonder why we do not model an election after each period. However, as will become clear, after the outcomes of first period policies are realized and revealed, the optimal policy for the second period is determined and will be chosen by any official—whether this is the previously elected incumbent or the winner of a new election. Thus, having an election after the first period would not change the results at all.

government; under centralization, a single election occurs nationwide to choose the central government. We assume elections pit the incumbent against a single alternative candidate. Candidates simultaneously commit to programs of policies for the two post-election periods. Second period policies may be conditioned on the outcomes of the first. (That is, for each district the relevant decisionmaker announces a policy for period 1 and a vector of policies for period 2 the elements of which correspond to each possible set of outcomes from policies chosen in the three districts in period 1.) To win, the incumbent in a local election must win a majority of votes in his locality; in a central election, the incumbent must win a majority of votes nationwide. Incumbents get a positive payoff from reelection and zero from losing, and so choose policies to maximize their probability of reelection.

To focus on just the effects of decentralization on experimentation, we assume the preferences of all voters within each locality (over outcomes in their own localities) are identical.

The representative voter in each locality votes for the incumbent if $u_1^I + \mathbf{d}u_2^I \geq u_1^C + \mathbf{d}u_2^C$,

where u_j^z is the voter's payoff in period $j \in \{1,2\}$ from the policy promised by candidate $z \in \{I,C\}$, where I stands for incumbent and C for challenger, and $0 < \mathbf{d}$ measures the weight voters place on second period payoffs relative to first period payoffs.¹⁴

We normalize so the per-period payoff of the status quo policy to the local representative voter is zero. The per-period payoff of the local representative voter is $-F_i$ if an experimental policy E_i fails and is S_i if it succeeds, where $F_i, S_i > 0$. Thus, choosing E_i yields the representative voter an expected utility of $qS_i - (1-q)F_i$ in one period. Note that the

For simplicity, we model a setup in which the results of experimentation are realized within the incumbent official's term.

¹⁴ One can think of \mathbf{d} as a discount factor, in which case $\mathbf{d} < 1$. However, in other interpretations it might be that $\mathbf{d} \geq 1$. For example, the second period might last much longer than the first. The analysis applies equally in both cases.

representative voter can be either risk-neutral or risk-averse in this formulation. Holding the underlying wealth in each state fixed, the more risk-averse is the voter, the smaller is S_i/F_i .¹⁵

Thus, the representative voter's total payoff is: zero if policy A is chosen in both periods, $(1 + d)S$ if the district finds a successful experiment in the first period and repeats it in the second, $-dF$ if A is chosen in the first period and a failed experiment in the second, and so on. We assume for now—and relax this later—that each of the l possible experiments has the same probability of success, q , and the same payoffs for “success” and “failure” (S and $-F$ respectively). So for now each experiment has the same expected value. We also assume initially—and relax the assumption later—that the outcomes of different experiments are independent.

At the beginning of the second period, all governments observe the outcomes of all first period experiments. Suppose one unit has discovered a successful policy. This unit will adopt the same policy in the second period, for a payoff of S . If another unit copies this policy in the second period, its payoff will be gS , where g measures the similarity of the units and therefore the extent to which policies that work well in one also succeed in another. The larger is g , the more one unit can benefit from others' discoveries. Alternatively, one could suppose the payoff to success is the same in all units, but the probability of success differs across units. If an experiment succeeds in unit m , we assume this raises the probability it will succeed in $j \neq m$ from q to $q' > q$.¹⁶ The payoff

¹⁵ Suppose f is the wealth cost of failure, s is the wealth benefit of success, and a is the wealth consequence of the status quo policy. Voter j is risk averse if $u_j(qs + (1 - q)(-f)) > qu_j(s) + (1 - q)u_j(-f) \equiv qS - (1 - q)F$. Given the normalization of the status quo's payoff to zero ($u_j(a) = 0$), an increase in risk aversion (i.e. an increase in the concavity of the payoff function) must either reduce S or increase F , or both, any of which would reduce S/F . This can be seen easily by graphing a concave function u that crosses zero at a . Holding the point $(a, 0)$ fixed, in order to make the graph more concave one must reduce $u_j(s) = S$ and/or reduce $u_j(-f) = -F$.

¹⁶ This assumption is fundamental to the whole “states as laboratories” way of thinking. If scientists did not believe that results of experiments conducted in one lab would generally hold in other labs, there would be little point in experimenting. It is also highly intuitive: if a policy “works” in one district, it would be odd to think that this renders it *less* likely to work in another. This is not to deny that local districts differ in many ways; one need only assume that states of the world in different units are positively correlated.

to j of choosing this experiment in the second period rises from $qS - (1-q)F$ to $q'S - (1-q')F$. This is analytically identical to the previous formulation (to see this, set $\mathbf{g} = q' - (1-q')F/S$). We assume that: $\max[0, q - (1-q)F/S] \leq \mathbf{g} \leq 1$, so an experiment that succeeded elsewhere is always preferred to the status quo or to just choosing an experiment at random.¹⁷

Under decentralization, the three local incumbents and their challengers choose their policy programs non-cooperatively and simultaneously. Under centralization, the central incumbent and her challenger do the same. We focus on the Nash equilibrium and compare the equilibrium number of experiments under centralization, decentralization, and the socially optimal plan.¹⁸

The first-best solution

Consider a social planner who maximizes the expected sum of the payoffs of all voters. Let k be the number of units in which experiments occur in the first period. Let $w \equiv \max\{0, qS - (1-q)F\}$.

The social planner's expected payoff is:

$$V = \begin{cases} 3\mathbf{d}w, & \text{if } k = 0; \\ qS - (1-q)F + \mathbf{d}[q(1+2\mathbf{g})S + 3(1-q)w], & \text{if } k = 1; \\ 2[qS - (1-q)F] + \mathbf{d}[(2+\mathbf{g})q^2S + 2q(1-q)(1+2\mathbf{g})S + 3(1-q)^2w], & \text{if } k = 2; \\ 3[qS - (1-q)F] + \mathbf{d}[3q^3S + 3q^2(1-q)(2+\mathbf{g})S + 3q(1-q)^2(1+2\mathbf{g})S + 3(1-q)^3w], & \text{if } k = 3. \end{cases}$$

Define $\mathbf{m} \equiv (1-q)F/(qS)$ as the risk ratio of experimentation. If $\mathbf{m} \leq 1$, all units will experiment in all periods until a successful policy is found, whether under centralization, decentralization, or a benevolent social planner. Thus, to focus on interesting cases, we suppose

¹⁷ In fact, we will focus on the case in which $q - (1-q)F/S < 0$, so \mathbf{g} is required to be in $[0,1]$. When $q - (1-q)F/S \geq 0$, the analysis becomes trivial since experimentation will take place in all units in the first period under both centralization and decentralization, as well as in the first best. Note that given $\mathbf{g} \leq 1$, $q' = (F + \mathbf{g}S)/(F + S) \leq 1$ and $q' > 0$. We consider a more general kind of heterogeneity (in which the payoff for success of a given experiment can vary more widely across units) in Section 3.

¹⁸ After first period outcomes are observed, the optimal second period policy is obvious to all. Each politician's program must offer the optimal second period plan or she will lose to her rival. Thus, time inconsistency issues do not arise.

that $m > 1$, or $(1 - q)F > qS$. This implies that $w = 0$, so no experimentation will occur in the second period if no successful policy is discovered in the first. Comparing the social planner's expected payoffs for different k , we find the first-best experimentation policy in period 1, k^* :

$$k^* = \begin{cases} 0, & \text{if } m > 1 + d(1 + 2g); \\ 1, & \text{if } 1 + d[1 + (2 - 3q)g] < m \leq 1 + d(1 + 2g); \\ 2, & \text{if } 1 + d[1 + (2 - 6q + 3q^2)g] < m \leq 1 + d[1 + (2 - 3q)g]; \\ 3, & \text{if } m \leq 1 + d[1 + (2 - 6q + 3q^2)g]. \end{cases}$$

Clearly, k^* is non-decreasing in d and non-increasing in m , which is quite intuitive. For relatively small q , k^* is non-decreasing in g , because the social benefit from an additional experiment is larger, the more applicable the knowledge acquired is to other units. However, for relatively large q , k^* is no longer monotonic in g . When the probability of success in any one experiment is very high, broader applicability of experimental discoveries encourages the planner to reduce his risk by experimenting in fewer units. Thus k^* may be decreasing in g .

Equilibrium under decentralization

To win election under decentralization, each local candidate must offer a policy plan that maximizes the expected payoff of the unit's citizens, taking into account the policies in other units. If a candidate does not do so, his opponent will propose the optimal policy and win. If both candidates propose policies with the same ex ante expected payoff for local citizens, each will be elected with 50 percent probability. Note that if there were only one unit, it would experiment in the first period if and only if $m \leq (1 + d)$. Table 1 shows the expected payoffs of a unit that experiments in the first period and of one that does not for each value of k .¹⁹

¹⁹ For instance, if one unit experiments, its expected first round payoff is $qS - (1 - q)F$ and its expected second round payoff is dqS (with probability q the experiment succeeds and is repeated in round two yielding dS ; with probability $(1 - q)$ it fails, and the unit chooses the status quo in round two, yielding 0). If

Table 1: Citizen payoffs under decentralization

	Experimenting unit's payoff	Non-experimenting unit's payoff
$k = 0$		0
$k = 1$	$qS - (1-q)F + d qS$	$d qgS$
$k = 2$	$qS - (1-q)F + d[qS + (1-q)qgS]$	$d(2q - q^2)gS$
$k = 3$	$qS - (1-q)F + d[qS + (1-q)(2q - q^2)gS]$	

This game has both a pure strategy equilibrium and a symmetric mixed-strategy equilibrium. We focus on the mixed strategy equilibrium here since it is the unique symmetric equilibrium. In the working paper version of the article, we show that similar and somewhat stronger results hold under the pure strategy equilibrium.

Let \mathbf{b} be the probability a given unit experiments in the first period in the mixed strategy equilibrium. It is easy to obtain the following result (all technical proofs are in the Appendix).

Lemma 1: *Under decentralization, in the mixed strategy equilibrium each locality chooses experimentation with probability*

$$\mathbf{b} = \begin{cases} 0, & \text{if } m > 1 + d; \\ \frac{1}{q} \left[1 - \left(1 - \frac{1 + d - m}{dg} \right)^{\frac{1}{2}} \right], & \text{if } 1 + d[1 - (2q - q^2)g] < m \leq 1 + d; \\ 1, & \text{if } m \leq 1 + d[1 - (2q - q^2)g]. \end{cases}$$

In the mixed strategy equilibrium, let k^d be the expected number of localities choosing experimentation in the first period, then $k^d = 3\mathbf{b}$. Intuitively, k^d is non-increasing in m . Moreover, k^d is non-increasing in g . As g rises, each unit has a stronger incentive to free ride on other units' experiments, thus reducing equilibrium experimentation.

two units experiment, then each has a probability q that its experiment succeeds and a probability $(1-q)q$ that its experiment fails but the other's succeeds. The first case yields a second round payoff of dS ; in the second case, the experimenter who failed will implement the other's successful experiment, yielding dgS .

Proposition 1: *Decentralization leads to less or equal experimentation as in the first best, i.e.,*

$k^d \leq k^*$, *except when* (i) $q \geq 0.687$ and $1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}[1 - \frac{4}{9}(3q - q^2)\mathbf{g}]$;

or (ii) $q \geq 0.825$ and $1 + \mathbf{d}[1 + (2 - 3q)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}[1 - \frac{1}{9}(6q - q^2)\mathbf{g}]$.

Proposition 1 demonstrates the *information externality effect*. Local policy experimentation generates information that benefits other localities. Under decentralization, local governments ignore such externalities and experiment less than is optimal, producing fewer discoveries on average than in the first best. This can be seen most clearly by comparing the condition for there to be at least some experimentation in the first best, $\mathbf{m} \leq 1 + \mathbf{d}(1 + 2\mathbf{g})$, with the condition for there to be at least some experimentation under decentralization, $\mathbf{m} \leq 1 + \mathbf{d}$. The former condition says it is optimal for at least one district to experiment if the expected benefit of success to *all three districts*, $(1 + \mathbf{d}(1 + 2\mathbf{g}))qS$, outweighs the expected cost of failure to the experimenting unit, $(1 - q)F$. The latter condition considers only the expected benefit of success to the experimenting unit, $(1 + \mathbf{d})qS$. The parameter \mathbf{g} , which measures the usefulness of information from one district to citizens in others, can serve as a proxy for the scope of informational externalities. Clearly the divergence from optimality—the difference between $1 + \mathbf{d}(1 + 2\mathbf{g})$ and $(1 + \mathbf{d})$ —increases with \mathbf{g} .

In the special cases in which decentralization leads to more experiments than in the first best, the experiments must be quite likely to succeed. Note that the one-period expected payoff from experimenting is still negative ($\mathbf{m} > 1$). Thus, in the case the success probability is very good, if one locality chooses to experiment in the first period, the others should wait and see the result of the first period experiment by that locality. However, in the symmetric mixed strategy equilibrium, local politicians do not coordinate perfectly and over-experiment relative to the first best. In the

(asymmetric) pure strategy equilibrium, such miscoordination does not occur and one can show that decentralization always leads to less or equal experiments than in the first best.

Equilibrium under centralization

Under centralization, the central candidates simultaneously propose two-period programs for each of the three units. The challenger wins if voters in at least two units prefer her policies.

Lemma 2: *Under centralization, the number of units chosen to experiment in the first period is 3 if and only if $\mathbf{m} \leq 1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}]$. Otherwise, there is no pure strategy equilibrium in the game of electoral competition between the incumbent and his challenger. In the symmetric mixed strategy equilibrium, the expected number of units chosen to experiment is given by:*

$$k^c = \begin{cases} 1, & \text{if } \mathbf{m} > 1 + \mathbf{d}[1 + (2 - 3q + q^2)\mathbf{g}]; \\ 2, & \text{if } 1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}[1 + (2 - 3q + q^2)\mathbf{g}]; \\ 3, & \text{if } \mathbf{m} \leq 1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}]. \end{cases}$$

Like k^* , k^c is non-decreasing in \mathbf{d} and non-increasing in \mathbf{m} ; it is non-decreasing in \mathbf{g} for relatively small q (specifically, $1 - 3q + q^2 > 0$) but is not monotonic in \mathbf{g} for large q . Like the social planner, central candidates choose policies that give voters the highest expected payoffs. However, unlike the social planner, candidates need to please only two of the three units. At times, therefore, they will sacrifice the citizens of one district to benefit those of the other two. This will not always result in the first best program, as the following result demonstrates.

Proposition 2: *Centralization leads to at least as much experimentation as in the first best, i.e., $k^c \geq k^*$, except when $1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}] < \mathbf{m} < 1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}]$ and $q < 0.5$.*

As Proposition 2 shows, in most cases centralization leads to at least as much experimentation as in the first best. The main reason is what we call the *risk-taking effect*. Since

candidates for central office need only win in a subset of districts, they use voters in other units as guinea pigs in the search for knowledge that will benefit their supporters elsewhere. The risk to victims of such experiments may be so great that not experimenting at all would be socially optimal. For instance, if $m > 1 + (1 + 2g)d$, experiments are so risky that the social planner would choose the status quo. Under centralization, the candidates completely write off the welfare of voters not needed for their coalition: in equilibrium, both mix with equal probability of choosing experiments in 0, 1, and 2 units, yielding one experiment on average.

However, Proposition 2 also reveals that under certain conditions centralization can lead to fewer experiments than in the first best. This is because of what we call the *risk-conserving effect*. Since central candidates ignore the welfare of those in the unit not needed for their coalition, they disregard the benefit of discoveries to those in the excluded unit. Suppose experiments have a relatively low probability of success ($q < 0.5$) but the payoff to success is large (F/S is small) so the risk ratio, m , is less than $1 + d[1 + (2 - 6q + 3q^2)g]$. From the social planner's perspective, all three units should experiment in the first period to fully exploit the positive externalities. However, if experiments are quite risky— $q < 0.5$ and $m > 1 + d[1 + (1 - 3q + q^2)g]$ —the candidates would want to reserve one unit and only experiment in the other two. Voters in the non-experimenting unit would bear no risk but still benefit from discoveries in the experimenting units, leaving them better off than if they had experimented. In equilibrium, both candidates mix with equal probability of choosing experiments in 1, 2, and 3 units, resulting in two experiments on average.

In sum, the fact that central candidates must win a majority nationwide implies three effects. The *information externalities effect* motivates central governments to experiment more than local ones would because they represent a broader set of constituents who have more to gain in the aggregate from any discovery. The *risk-taking effect* motivates central governments to experiment in some units even when the chance of success is low because they can afford to write off the

voters in the experimenting districts and their supporters elsewhere will benefit from discoveries.²⁰ However, in other circumstances the *risk-conserving effect* causes central governments to experiment less than is socially optimal (although usually at least as much as under decentralization: see below) because they do not consider the benefit of discoveries to voters not in their coalition.

2 Comparing centralization and decentralization

Comparing centralization and decentralization, we have the following result:

Proposition 3: *Centralization leads to at least as much experimentation as decentralization, i.e.,*

$$k^c \geq k^d, \text{ except when } 1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}] < \mathbf{m} < 1 + \mathbf{d}\left[1 - \frac{4}{9}(3q - q^2)\mathbf{g}\right] \text{ and } q > 0.83.$$

Thus, in most cases, centralization leads to at least as many experiments as occur in the mixed strategy equilibrium under decentralization.²¹ The reason is clear from Propositions 1 and 2. Local officials under decentralization ignore the information externality and experiment less than is optimal. Motivated by the risk-taking effect, central officials experiment *more* than is optimal. At times (when experiments are rather risky), the risk-conserving effect can overwhelm the risk-taking effect, reducing experimentation under centralization to sub-optimal levels. But centralization still yields more experiments than decentralization in such cases. Electoral competition causes central candidates to ignore externalities that benefit those outside their coalition—but not those that benefit supporters. Under decentralization, candidates ignore all positive externalities. Only in a class of cases in which the experiments are very likely to succeed does decentralization lead to more experiments than centralization. Intuitively, when experiments are very likely to succeed,

²⁰ The notion that in a decentralized state a majority of subunits may cooperate to exploit a minority is, of course, a familiar one. It has been modeled recently by Crémer and Palfrey (2006) and Hafer and Landa (2007) in the context of fiscal redistribution. Since Riker (1962), models of legislative bargaining have also often derived results in which those in the minimum winning coalition exploit those outside it.

²¹ If the pure strategy equilibrium occurs under decentralization, centralization leads to at least as many experiments as decentralization for all parameter values.

local politicians mis-coordinate by over-experimenting. On the other hand, the risk-conserving effect under centralization is strong (i.e., the one-period expected payoff is still negative and voters in one unit are very eager to observe others' experiments rather than experiment themselves), so decentralization may lead to more experimentation than centralization.

The welfare comparison between centralization and decentralization is ambiguous. But, contrary to the common intuition, the electoral effects of political decentralization tend to produce too little experimentation under decentralization and too much under centralization. Insofar as the systems yield different results, centralization is preferable when experiments are valuable and broadly applicable, while decentralization may be better when experiments are risky and the main concern is to prevent citizens being used as guinea pigs.

3 Extensions

A large number of localities

Increasing the number of local districts has two effects that should tend to produce more experiments under centralization and fewer under decentralization. First, the risk-taking effect gets stronger. The central government need only win support of voters in a majority of units. As the number of units increases, the importance of each to the center's coalition declines. The center will be willing to risk failure in a larger number of districts in search of successful policies to implement in the others. By contrast, under decentralization each local government must win a majority within its own jurisdiction, and this does not change with the number of localities. Second, the more districts there are, the stronger is the incentive for individual localities under decentralization to wait and free-ride on others' discoveries.

Specifically, suppose there are $N = 2n + 1$ localities, where $n \geq 1$. Under centralization, it is easy to see that in equilibrium the number of units chosen to experiment in the first period, k^c , must be greater than or equal to n . If a candidate proposed to experiment in $m < n$ districts, his

rival could win for sure by proposing experiments in n units, as all the $n+1$ others would be better off. Under decentralization, we can easily see that the free-riding problem worsens as N increases in the mixed strategy equilibrium. The mixed strategy equilibrium with N localities can be characterized as follows:

Lemma 3: *Under decentralization, in the mixed strategy equilibrium each locality chooses experimentation with probability*

$$\mathbf{b} = \begin{cases} 0, & \text{if } m > 1 + d; \\ \frac{1}{q} \left[1 - \left(1 - \frac{1 + d - m}{dg} \right)^{N-1} \right], & \text{if } 1 + d[1 - (1 - (1 - q)^{N-1})g] < m \leq 1 + d; \\ 1, & \text{if } m \leq 1 + d[1 - (1 - (1 - q)^{N-1})g]. \end{cases}$$

In the mixed strategy equilibrium, the expected number of localities choosing experimentation is $k^d = N\mathbf{b}$. As N rises, the free-riding problem gets worse. In the interesting case in which $\mathbf{b} \in (0,1)$, we can prove the following result.

Proposition 4: *Under decentralization, in the (strictly) mixed strategy equilibrium, the probability of each locality experimenting goes to zero as N increases, and the expected number of experiments converges to a constant independent of N . In contrast, under centralization, the number of experiments increases with N without bound.*

Correlated experiments

So far, we have assumed politicians choose from a large set of policy experiments, each of which has the same benefits if successful, costs if unsuccessful, and probability of success. We also assumed the results of all experiments are independent. But often these assumptions will not hold. We first examine what happens if experiments' results are not independent. For instance, some experiments involve choosing different points on a scale. Suppose the speed limit on highways is 65 miles per hour. One region might reduce its limit to 55 miles per hour to see how

this affects traffic accidents, another to 45 miles per hour, and so on. The results would likely be correlated, though less than perfectly.

To analyze correlated experiments, we consider the simplest case of perfect correlation, in which there is only one available experiment. All units that experiment implement this policy and have the same outcome. As before, we continue to assume $m > 1$ to focus on the interesting case. All localities implement the experiment in the second period if and only if some unit chose it in the first period and it was successful. In this case, if at least one unit experiments in the first period, then a unit's expected payoff is $qS - (1-q)F + d q S$ if it experiments and $d q g S$ if it does not, regardless of the number experimenting. It is easy to verify that the socially optimal policy is $k^* = 0$ if $m > 1 + d(1 + 2g)$; $k^* = 1$ if $1 + d(1 - g) < m \leq 1 + d(1 + 2g)$; and $k^* = 3$ if $m \leq 1 + d(1 - g)$. Under centralization, both candidates will choose the following policy: $k^c = 1$ if $1 + d(1 - g) < m$; and $k^c = 3$ if $m \leq 1 + d(1 - g)$. Under decentralization, $k^d = 0$ if $m > 1 + d$; $k^d = 1$ if $1 + d(1 - g) < m \leq 1 + d$; and $k^d = 3$ if $m \leq 1 + d(1 - g)$.

When $m \leq 1 + d(1 - g)$, $k^d = k^* = k^c = 3$. When $1 + d(1 - g) < m \leq 1 + d$, $k^d = k^* = k^c = 1$. In these two cases, decentralization and centralization both yield the first best outcome. When $1 + d < m \leq 1 + d(1 + 2g)$, $k^d = 0 < k^* = k^c = 1$: there is less experimentation under decentralization than under centralization or in the first best, because of free riding. When $1 + d(1 + 2g) < m$, $k^d = k^* = 0 < k^c = 1$: centralization yields more experiments than under decentralization or in the first best, because of the risk-taking effect. Similar conclusions should hold for highly, but imperfectly, correlated experiments. In sum, for perfectly correlated experiments, we get results qualitatively similar to those derived in previous sections.

Proposition 5: *For perfectly correlated experiments, the level of experimentation under centralization is always at least as great as—and sometimes greater than—under decentralization.*

When experiments are not perfectly correlated, centralization may have another benefit. Consider the example of choosing different speed limits to explore the effect of this on traffic accidents. The amount of information that results from a given number of experiments will depend on how the different treatments are selected. A central government would probably choose speed limits spaced across the relevant range on the scale, making it easier to discern the relationship. Under decentralization, uncoordinated local units may duplicate each other's choices or bunch near certain values, making it harder to estimate the relationship with confidence. This *calibration effect* will tend to render correlated experiments under centralization more informative.

Heterogeneous experiments

We now relax the assumption that all policy experiments have the same probability of success. Suppose there is one experiment E_0 with success probability q and many others with success probability $\hat{q} < q$ (call these “lower-odds experiments”).²² All experiments are statistically independent and still have the same payoffs for success and failure, respectively S and $-F$. To focus on the most interesting case, we assume $\mathbf{m} = (1 - q)F/qS \in (1, 1 + \mathbf{d})$, so at least one unit experiments in the first period under decentralization, centralization, and in the first best. Clearly, since E_0 is better than lower-odds experiments, it will always be chosen if there is any experimentation. But will other districts experiment, and if so what experiments will they choose?

Because $\mathbf{m} > 1$, neither the social planner nor central politicians will choose E_0 in more than one unit. Thus, if they experiment in more than one unit in the first period, they will implement lower-odds experiments. The analysis parallels that in the sections on “The first-best solution” and “Equilibrium under centralization,” yielding a result similar to Proposition 2. Under decentralization, however, multiple units may simultaneously choose E_0 , duplicating each other's

²² As $\hat{q} \rightarrow 0$, this case converges to that of perfectly correlated experiments considered in the previous subsection.

effort and taking unnecessary risk. Without loss of generality, suppose unit 1 chooses E_0 . Unit 2's payoff is: $V^E = \hat{q}S - (1 - \hat{q})F + \mathbf{d}[\hat{q}S + (1 - \hat{q})\mathbf{g}S]$ if it chooses a lower-odds experiment; $V^{E_0} = qS - (1 - q)F + \mathbf{d}qS$ if it adopts E_0 ; and $V^{NE} = \mathbf{d}q\mathbf{g}S$ if it chooses A. Thus, when $V^{E_0} > V^E$ and $V^{E_0} > V^{NE}$ (which can easily be true if $\mathbf{m} < 1 + \mathbf{d}(1 - \mathbf{g})$ and \hat{q} is small relative to q , or \mathbf{g} is relatively small), units 2 and 3 also choose E_0 in the first period. Lower-odds experiments, even with reasonably good chances of success, are never tried.

In short, when experiments are heterogeneous, centralization has an additional advantage. Under decentralization, governments may neglect lower-odds experiments that have reasonably high probabilities of success because they fail to coordinate and all pick the same experiment. Under centralization, the central government implements different experiments in different localities so as to maximize the chance of success.

Heterogeneous districts

Suppose now that voters in different districts have different preferences over the experimental policies available. For instance, voters in left-wing districts might derive higher payoffs from experimenting with universal health insurance than right-wing units; the latter might have higher payoffs than left-wing districts from experimenting with harsh criminal penalties. We could assume that a given experiment has the same F but a different value of S in different units. Clearly, local governments under decentralization will prioritize experiments that, for a given chance of success, best match the tastes of local voters (i.e. have the highest local S).

Under centralization, the candidates would offer districts in their support coalition the experiments that local voters most preferred. But for units not in their coalition, the candidates would choose the experiments that (for a given \mathbf{g}) have the highest value of S for voters in the support coalition. Not only may local voters in units outside the coalition be subjected to greater risk than they would chose for themselves, they may end up guinea pigs in their least favorite

experiments. In some cases, this could increase the parameter range in which decentralization produces greater social surplus precisely by limiting central experimentation. Such preference heterogeneity may also reduce the incentive for a given local government to free-ride, since the experiments other districts choose may be ones the given unit's voter's dislike.²³

Negative externalities

One of the main advantages of centralization is that it motivates officials to at least partially internalize the positive informational externalities of local experiments. But externalities associated with experiments can also be negative. Suppose local governments reduce pollution controls to see if this attracts investment or cut welfare payments to see if this motivates the poor to move.

Clearly, a central government that internalizes such externalities will enact fewer such experiments than local governments that can export the costs. Since central candidates need votes from more than one unit, they will care about costs that cross local borders. Equally clearly, if the external costs of such experiments are high, preventing them will increase social welfare.

One kind of interregional externality merits separate consideration. Suppose voters care what policies are enacted not just in their own unit but in other districts too. In the US, for instance, many voters feel strongly about whether abortion should be legal in other states besides their own.²⁴ Voters might care about policy in other units because they anticipate moving to them. Or they might simply wish to impose their values on others. If such issues inform voting, the central

²³ If experiments have a direct cost, it might seem that the central government would be less likely to choose certain experiments than a locality that had a strong preference for those experimental policies. However, if the local government under decentralization would choose an experiment despite the direct cost, the central government could recover the cost if it chose this experiment simply by taxing the locality—the local population would be willing to pay. So such costs should not affect the calculus. In general, we have abstracted from questions of interregional redistribution since these would distract from the issue of experimentation and are analytically distinct from it. If problems of fiscal interdependence between center and localities lead to inefficiencies, this will be the case whether or not policies are experimental.

²⁴ Calabresi and Melamed (1972, pp.1111-12, quoted in Gillette 1997) call such externalities “moralisms,” or external costs that “do not lend themselves to collective measurement which is acceptably objective and nonarbitrary.”

government will take them into account; it may not choose locally popular policies that alienate voters elsewhere. Banning abortion in Alabama might win central candidates votes there but lose them even more in California.²⁵ This might reduce the number of experiments under centralization.

Although such ideological spillovers may be important, they may not imply a great innovation advantage for decentralization in practice. First, institutions guaranteeing local autonomy must be quite robust to resist the pressure of national public opinion. Even in countries with very decentralized constitutions, such as the US, central politicians and judges often intervene to overrule subnational governments whose policies offend national opinion. The US Supreme Court uses the Commerce Clause, the 8th Amendment, and Section 5 of the 14th Amendment to invalidate state laws on various grounds, often explicitly rejecting the “states as laboratories” argument—as, of course, the majority did in *New State Ice Co. vs. Liebmann*. To overturn state penal policies, the Court has invoked “evolving standards,” reflecting its reading of nationwide majority opinion. So ideological spillovers may curb unpopular experiments even in decentralized systems. Second, if a policy is abhorrent to the nationwide majority, the potential for it to spread, if successful, is limited. Suppose a US city were able to legalize heroin and found that drug overdoses fell; it seems doubtful many other cities would follow. The experiments centralization prevents under this argument are precisely those likely to generate few benefits nationwide.

And third, in a decentralized order with three or more tiers, ideologically motivated regional governments or voters may limit experimentation in their jurisdictions, banning even experiments the nationwide majority might have permitted. In the 1990s, some municipalities in Colorado enacted ordinances banning anti-gay discrimination; these were overturned by a state-wide voter initiative (“Amendment 2”), but then reinstated by the US Supreme Court.²⁶ For

²⁵ Central governments might also be more vulnerable to pressures from national interest groups such as labor unions or business confederations, which want uniform policies for their members nationwide. This might lead central governments to ignore local variation. However, national interest groups could also lobby local governments in a decentralized system. Indeed, their threats to punish individual local units with boycotts, etc., might be more credible than threats to stage nationwide strikes or other actions.

²⁶ “*Romer vs. Evans*,” 116 S. Ct. 1620 (1996). See the discussion in Gillette (1997).

decentralization to protect local experimentation, not only must the autonomy of local units be extremely securely entrenched, autonomy must be entrenched in the smallest local units.

4 Discussion

An additional *communication effect*, not easily incorporated into the model, suggests another advantage of centralization. Information dissemination likely exhibits economies of scale. It is cheaper for a central government to consolidate reports of local experiments and publicize these nationwide than for each local government to survey all the others and do its own analysis. The central government might still perform this service even under decentralization, but it will do so for sure if it has authorized the experiments itself. For this reason, John Stuart Mill argued for a vigorous central government in *On Representative Government*: “The principal business of the central authority should be to give instruction, of the local authority to apply it. Power may be localized, but knowledge, to be most useful, must be centralized; there must be somewhere a focus at which all its scattered rays are collected, that the broken and coloured lights which exist elsewhere may find there what is necessary to complete and purify them” (Mill 1991, p.424).

Obtaining information about local experiments may not be just a technical matter. Sometimes, voters will not observe policies, but only the resulting performance. Suppose a local government designs a procedure for tracking bureaucratic waste that reduces costs substantially. In theory, the procedure could be introduced in other districts too. But suppose the local incumbent under decentralization can keep elements of the procedure secret (perhaps it requires particular software). Local voters may evaluate the performance of their local governments *relative* to that in neighboring districts using yardstick competition (Shleifer 1985, Besley and Case 1995). If so, the innovative local government will not want to share details of its discovery because others’ use of it will erode the first unit’s relative performance. Under centralization, no such problem arises (unless the government also uses yardstick competition to reward agents and cannot get them to reveal

policy details). So the fruits of experimentation may be shared more rapidly under centralization for strategic as well as technical reasons.²⁷

Although the claim that decentralization's main benefit is to limit policy experiments contradicts a common intuition, it is not as unprecedented as it might at first seem. One of the oldest arguments for political decentralization is that it can restrain abuses by a central government that is *too* willing to force risky policies onto local communities. The Stamp Act, which helped set off the American Revolutionary War, was attacked by colonists as a “dangerous innovation,” imposed by a distant, irresponsible central authority (McDonald 1962). The U.S. constitution, one of the most decentralized, was intended, according to Chief Justice William Taft writing in 1921 to “prevent experimentation with the fundamental rights of the individual.”²⁸

Indeed, there is reason to think that the decentralized political structure of the US at times retards—rather than stimulates—the adoption of novel policies. It is sometimes suggested that the US is one of the world's leading policy innovators. As evidence of this, scholars often point to cases in which state-level experiments inspired later federal programs. In the 1920s and 1930s, state innovations in social policy and economic regulation laid the ground for much of President Roosevelt's New Deal (Patterson 1969, Morehouse and Jewell 2004). However, the relevant question is not whether local innovations preceded central legislation in the US, but whether a decentralized country like the US was quicker to innovate than its more centralized counterparts. In fact, the US was one of the slowest developed countries to adopt social legislation. Among the 30 current OECD members, the US was 23rd to pass national legislation on old age pensions, disability benefits or insurance, or benefits for surviving dependents of the insured. It fell behind Slovakia, Ireland, Iceland, Italy, and Spain, among other countries. The US was the last of the 30 to pass

²⁷ Another issue concerns policy experiments—in foreign policy, say, or monetary management—that are inherently national in scope. If decentralization increases the number of institutional veto points, as in some federal structures, this may make it harder to obtain the agreement necessary to enact such experiments.

²⁸ *Truax v. Corrigan* (1921, 257 U.S. 312). Taft was rebutting an early argument for states as laboratories made by Justice Oliver Wendell Holmes.

national legislation on sickness and maternity benefits. These comparisons relate to national legislation, but even if one focused on the first state laws on pensions, disability or survivor insurance around 1920, the US would still come only about 18th out of 30.²⁹

Our point here is not that, in creating early social security programs, centralized European democracies implemented local experiments in the manner modeled in this paper. On the contrary, almost everywhere new social policies in the late 19th and early 20th centuries appear to have built on experiments by voluntary organizations or local governments. Rather, the fact that many centralized European democracies adopted national schemes faster than did the US probably reflects the greater speed with which the centralized democracies were able to take up and extend local policies that seemed promising. Decentralized political institutions, by creating more vetoes over central policy, should tend to slow the centrally mandated spread of local discoveries.

Indeed, policy innovations may tend to come from local governments in decentralized countries precisely because local governments in such countries block central attempts to innovate. In Canada, the province of Saskatchewan has been seen as a “crucial incubator” of social policy, paving the way for national hospital insurance in 1957 and national health insurance in 1966 (Hacker 1998, pp.72, 104; Gray 1991). But Saskatchewan’s experiment in the late 1940s came after the federal government had been trying to introduce health insurance for decades. Ottawa attempted to do so during the Great Depression but was blocked by the country’s top constitutional tribunal, which ruled that health insurance fell within the provinces’ jurisdiction (Hacker 1998, p.97). The Liberal Mackenzie King government tried again in 1945, but could not get the necessary provincial agreement. Only after several provinces introduced their own programs, following

²⁹ These data come from *Social Security Programs Throughout the World, 2004*, compiled by the International Social Security Association, in Geneva, Switzerland, downloaded from the US Social Security Administration, at www.socialsecurity.gov, June 23, 2006. The US can claim with more justification to have been a leader in reforms to introduce work requirements for welfare recipients. And state initiatives in the 1980s and 1990s did lay the ground for the 1996 federal welfare reform. Still, Germany and Denmark had begun experimenting with such schemes even earlier, in the 1960s and 1970s, and other countries (e.g. France) introduced reforms around the same time as the US (Lodemel and Trickey 2000). Britain’s 1990s reforms to introduce work requirements again show that highly centralized countries can and do adopt local experiments. The 1996 Project Work program was piloted in Hull and Medway/Maidstone, followed by 29 other areas (Ritchie and Legard 1997).

Saskatchewan's lead, was Ottawa allowed to superimpose a common framework. If provincial governments in federations block all central attempts to innovate, then of course all innovations—if they occur—must start in the provinces. But this does not mean federalism stimulates innovation. Like the US, Canada was a laggard not a leader in introducing social policies. Its national hospital insurance in 1957 came extremely late compared to other developed countries. The median date of the first sickness or maternity benefits program among current OECD members was 1921.

5 Conclusion

If central governments can enact geographically differentiated policies, then whether centralization or decentralization leads to greater policy experimentation must depend on the incentives for officials under the two systems. Modeling the electoral logic of centralized and decentralized democracies, we did not find support for the common belief that political decentralization stimulates innovation. In fact, compared to the first best outcome, decentralization tended to produce too little experimentation, while centralization usually led to too much—and usually to more than decentralization. The experimentation gap in favor of centralization increases with the number of local districts and, in most cases, with the cross-regional applicability of discoveries. Centralization tends to produce a larger total payoff to citizens when experimentation is valuable. By contrast, decentralization can be preferable precisely because it restrains local experiments, when such experiments have low value and narrow applicability or when voters are very risk averse. If experiments have large negative externalities, decentralization may generate more experiments—but ones that reduce social welfare. Centralization may also lead to better design and coordination of experimental programs as well as dissemination of their results.

References

- Alesina, Alberto, Nouriel Roubini, and Gerald Cohen. 1997. *Political Cycles and the Macroeconomy*. Cambridge, MA: MIT Press.
- Althouse, Ann. 2004. "Vanguard States, Laggard States: Federalism and Constitutional Rights." *University of Pennsylvania Law Review* 152: 1745-96.
- Barro, Robert. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14(March): 19-42.
- Besley, Timothy and Anne Case. 1995. "Incumbent Behavior: Vote Seeking, Tax Setting, and Yardstick Competition." *American Economic Review* 85(March): 25-45.
- Besley, Timothy and Stephen Coate. 2003. "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach." *Journal of Public Economics* 87(December): 2611-37.
- Bilinsky, Yaroslav. 1968. "Education of the Non-Russian Peoples in the USSR, 1917-1967: An Essay." *Slavic Review* 27(September): 411-37.
- Bordignon, Massimo, Paolo Manasse, and Guido Tabellini. 2001. "Optimal Regional Redistribution Under Asymmetric Information." *American Economic Review* 91(June): 709-23.
- Brennan, Geoffrey and James M. Buchanan. 1980. *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. New York: Cambridge University Press.
- Breton, Albert. 1996. *Competitive Governments: An Economic Theory of Politics and Public Finance*. Cambridge: Cambridge University Press.
- Bryce, James. 1888. *The American Commonwealth*. Vol. I, New York: MacMillan and Co.
- Calabresi, Guido and A. Douglas Melamed. 1972. "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral." *Harvard Law Review* 85(April): 1089-1128.
- Crémer, Jacques and Thomas R. Palfrey. 2006. "An equilibrium voting model of federal standards with externalities." *Journal of Public Economics* 90(November): 2091-2106.
- Ekedahl, Carolyn and Melvin A. Goodman. 1997. *The Wars of Eduard Shevardnadze*. University Park, PA: Penn State University Press.
- Eley, Susan, Kathryn Gallop, Gill McIvor, Kerry Morgan, Rowdy Yates. 2002. *Drug Treatment and Testing Orders: Evaluation of the Scottish Pilots*. Edinburgh: Scottish Executive Central Research Unit, at <http://www.scotland.gov.uk/cru/kd01/green/dtts.pdf>.
- Fair, Ray. 1978. "The Effect of Economic Events on Votes for President." *Review of Economics and Statistics* 60: 159-72.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 5-26.

- Gillette, Clayton P. 1997. "The Allocation of Government Authority: The Exercise of Trumps by Decentralized Governments." *Virginia Law Review* 83(October): 1347-1418.
- Gray, Gwendolyn. 1991. *Federalism and Health Policy: The Development of Health Systems in Canada and Australia*. Toronto: University of Toronto Press.
- Greve, Michael S. 2001. "Laboratories of Democracy: Anatomy of a Metaphor." *Federalist Outlook*, 6(May).
- Hacker, Jacob S. 1998. "The Historical Logic of National Health Insurance: Structure and Sequence in the Development of British, Canadian, and US Medical Policy." *Studies in American Political Development* 12(Spring): 57-130.
- Hafer, Catherine and Dimitri Landa. 2007. "Public Goods in Federal Systems." *Quarterly Journal of Political Science* 2: 253-275.
- Jowell, Roger. *Trying it Out: The Role of 'Pilots' in Policymaking*. London: Government Chief Social Researcher's Office, 2003.
- Kollman, Ken, John H. Miller and Scott E. Page. 2000. "Decentralization and the Search for Policy Solutions." *Journal of Law, Economics, and Organization* 16(1): 102-28.
- Kramer, Gerald. 1971. "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964." *American Political Science Review* 65: 131-43.
- Laski, Harold J. 1921. *The Foundations of Sovereignty and Other Essays*. New York: Harcourt, Brace and Co.
- Lockwood, Ben. 2002. "Distributive Politics and the Benefits of Decentralization," *Review of Economic Studies*. 69: 313-38.
- Lodemel, Ivar and Heather Trickey. 2000. "A New Contract for Social Assistance," in Lodemel and Trickey, eds., *An Offer You Can't Refuse: Workfare in International Perspective*. Bristol: Policy Press, pp.1-40.
- MacLennan, M.C. 1965. "Regional Planning in France." *Journal of Industrial Economics* 13: 62-75.
- McDonald, Forrest, ed. 1962. *Empire and Nation: John Dickinson, Letters from a Farmer in Pennsylvania and Richard Henry Lee, Letters from the Federal Farmer*. Indianapolis: Liberty Fund, 1999, second edition.
- Mill, John Stuart. 1991. *John Stuart Mill: On Liberty and Other Essays*. ed. John Gray, New York: Oxford University Press.
- Morehouse, Sarah M. and Malcolm E. Jewell. 2004. "States as Laboratories: A Reprise." *Annual Review of Political Science* 7: 177-203.
- Oates, Wallace. 1972. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.

-----, 1999. "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37(September): 120-49.

Patterson, James T. 1969. *The New Deal and the States: Federalism in Transition*. Princeton, NJ: Princeton University Press.

Riker, William H. 1962. *The Theory of Political Coalitions*. New Haven: Yale University Press.

Ritchie, Jane and Robin Legard. 1997. *The First Project Work Pilots: A Qualitative Evaluation*. London: UK Department for Education and Employment.

Rose-Ackerman, Susan. 1980. "Risk Taking and Reelection: Does Federalism Promote Innovation?" *Journal of Legal Studies* 9: 593-616.

Sah, Raaj and Joseph E. Stiglitz. 1986. "The Architecture of Economic Systems: Hierarchies and Polyarchies." *American Economic Review* 76(September): 716-27.

Seabright, Paul. 1996. "Accountability and decentralisation in government: An incomplete contracts model." *European Economic Review* 40: 61-89.

Shleifer, Andrei 1985. "A Theory of Yardstick Competition." *Rand Journal of Economics* 16(Autumn): 319-27

Strumpf, Koleman. 2002. "Does Government Decentralization Increase Policy Innovation." *Journal of Public Economic Theory* 4(2): 207-41.

UK Office of the Deputy Prime Minister, *Evaluation of a Pilot Seller's Information Pack: The Briston Scheme, Final Report*. London: Office of the Deputy Prime Minister, 2002.

at

http://www.odpm.gov.uk/stellent/groups/odpm_housing/documents/downloadable/odpm_house_601730.pdf

Appendix

Proof of Lemma 1: This result is a special case of Lemma 3 when $N = 3$. See the proof of Lemma 3 below. **Q.E.D.**

Proof of Proposition 1: (i) If $2 - 6q + 3q^2 \geq 0$ or $q \leq 0.37$, then $k^* = 3$ for all

$$\mathbf{m} \leq 1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}], \text{ while } k^d = 3\mathbf{b} \leq 3. \text{ For } \mathbf{m} > 1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}] \geq 1 + \mathbf{d},$$

we have $k^d = 0$, while $k^* = 2, 1, \text{ or } 0$. Thus, $k^* \geq k^d$ for all \mathbf{m} when $q \leq 0.37$.

(ii) Suppose $2 - 6q + 3q^2 < 0$ and $2 - 3q > 0$, or $0.37 < q < 2/3$. It can be verified that that

$$\frac{3}{q} \left[1 - \left(1 - \frac{1 + \mathbf{d} - \mathbf{m}}{\mathbf{d}\mathbf{g}} \right)^2 \right] < 2 \text{ when } \mathbf{m} > 1 + \mathbf{d} \left[1 - \frac{4}{9}(3q - q^2)\mathbf{g} \right]. \text{ For } q < 2/3, \text{ one can show that}$$

$$2 - 6q + 3q^2 > -\frac{4}{9}(3q - q^2). \text{ Thus, for } \mathbf{m} \leq 1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}], k^* = 3 \text{ and } k^d = 3\mathbf{b} \leq 3;$$

for $1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}[1 + (2 - 3q)\mathbf{g}]$, $k^* = 2$ and $k^d < 2$; and for

$$\mathbf{m} > 1 + \mathbf{d}[1 + (2 - 3q)\mathbf{g}], k^* = 1 \text{ or } 0 \text{ while } k^d = 0. \text{ Therefore, in this case } k^* \geq k^d \text{ for all } \mathbf{m}.$$

(iii) Suppose $2 - 3q \leq 0$ or $q \geq 2/3$. It can be verified that if $q < 0.687$, then

$$2 - 6q + 3q^2 > -\frac{4}{9}(3q - q^2), \text{ and the comparison of } k^* \text{ and } k^d \text{ is exactly as in (ii), that is,}$$

$$k^* \geq k^d. \text{ If } q \geq 0.687 \text{ and } 1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d} \left[1 - \frac{4}{9}(3q - q^2)\mathbf{g} \right], \text{ then}$$

$$k^* = 2 < k^d. \text{ Moreover, if } q \geq 0.825 \text{ and } 1 + \mathbf{d}[1 + (2 - 3q)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d} \left[1 - \frac{1}{9}(6q - q^2)\mathbf{g} \right], \text{ then}$$

$$k^* = 1 < k^d. \text{ For all other cases, we have } k^* \geq k^d. \quad \mathbf{Q.E.D.}$$

Proof of Lemma 2: First note that $k = 1$ always beats $k = 0$, in the sense that a player choosing $k = 1$ will win the election for sure against the other player choosing $k = 0$. This is because when $k = 1$ the two non-experimenting units benefit from the other unit experimenting. Similarly, $k = 2$

always beats $k = 1$, and $k = 3$ always beats $k = 2$. Thus, in a pure strategy equilibrium, no player will choose $k \in \{0, 1, 2\}$. If one does, his opponent can win for sure by proposing to change one unit's action from non-experimenting to experimenting while keeping the other two's actions unchanged, leaving the other two better off.

So if a pure strategy equilibrium exists, it must be that both the incumbent and his challenger propose experimentation in all units. For this to be an equilibrium, it must be that $qS - (1-q)F + d[qS + (1-q)(2q - q^2)gS] \geq dgqS$ so that no player wants to deviate to $k = 1$. If this is the case, then no player wants to deviate to either $k = 0$ or $k = 2$. Thus, $k = 3$ is a dominant strategy. Therefore, when $m \leq 1 + d[1 + (1 - 3q + q^2)g]$, the only equilibrium is that both will propose $k = 3$.

Now consider the case in which $m > 1 + d[1 + (1 - 3q + q^2)g]$. If in addition $m \leq 1 + d[1 + (1 - q)g]$, then $k = 2$ "beats" $k = 0$ since the two units chosen to experiment in $k = 2$ are better off than in $k = 0$. This also implies that $k = 3$ "beats" $k = 0$. Then, $k = 0$ is a dominated strategy. The whole game can be represented in the following matrix form, where L is the payoff from winning the election and $-L$ is the payoff from losing the election.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$
$k = 0$	0, 0	-L, L	-L, L	-L, L
$k = 1$	L, -L	0, 0	-L, L	L, -L
$k = 2$	L, -L	L, -L	0, 0	-L, L
$k = 3$	L, -L	-L, L	L, -L	0, 0

After deleting the dominated strategy $k = 0$, the reduced game is a standard "paper-scissors-rock" game. This game has a unique equilibrium in which both players randomize over the three strategies with equal probabilities. Therefore, the expected number of units chosen to experiment in this case is 2.

Next suppose $1 + d[1 + (1 - q)g] < m \leq 1 + d[1 + (2 - 3q + q^2)g]$. Then $k = 0$ beats $k = 2$

but is beaten by $k = 3$. In this case the game can be represented as

	$k = 0$	$k = 1$	$k = 2$	$k = 3$
$k = 0$	0, 0	-L, L	L, -L	-L, L
$k = 1$	L, -L	0, 0	-L, L	L, -L
$k = 2$	-L, L	L, -L	0, 0	-L, L
$k = 3$	L, -L	-L, L	L, -L	0, 0

Again $k = 0$ is a dominated strategy (by $k = 3$). So as in the previous case, the expected number of units chosen to experiment in the mixed strategy equilibrium is 2.

Finally, suppose $m > 1 + d[1 + (2 - 3q + q^2)g]$. Then $k = 0$ beats $k = 2$ and $k = 3$. In this case the game can be represented as

	$k = 0$	$k = 1$	$k = 2$	$k = 3$
$k = 0$	0, 0	-L, L	L, -L	L, -L
$k = 1$	L, -L	0, 0	-L, L	L, -L
$k = 2$	-L, L	L, -L	0, 0	-L, L
$k = 3$	-L, L	-L, L	L, -L	0, 0

Now $k = 3$ becomes a dominated strategy (by $k = 0$). After deleting $k = 3$, the reduced game has a symmetric mixed strategy equilibrium in which both players randomize over the three remaining strategies with equal probability. The expected number of units chosen to experiment in this equilibrium is 1. **Q.E.D.**

Proof of Proposition 2: If $q \geq 0.5$, then $1 - 3q + q^2 \geq 2 - 6q + 3q^2$. Thus, when $k^c = 3$,

$k^* = 3$ or 2; when $k^c = 2$, $k^* = 2$ or 1; and when $k^c = 1$, $k^* = 1$ or 0. Therefore, when $q \geq 0.5$,

$k^c \geq k^*$. When $1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}] < \mathbf{m} < 1 + \mathbf{d}[1 + (2 - 6q + 3q^2)\mathbf{g}]$ and $q < 0.5$,

$k^c = 2 < k^* = 3$. In all other cases, we still have $k^c \geq k^*$.

Q.E.D.

Proof of Proposition 3: Clearly, $k^c > k^d$ when $\mathbf{m} > 1 + \mathbf{d}$ since $k^d = 0$ and $k^c = 1$ or 2. Also,

$k^c = k^d = 3$ when $\mathbf{m} \leq 1 + \mathbf{d}[1 - (2q - q^2)\mathbf{g}]$.

(i) If $1 - 3q + q^2 \geq 0$, then $k^c = 3$ for all $1 + \mathbf{d}[1 - (2q - q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}]$,

while $k^d = 3\mathbf{b} < 3$. Thus, $k^c \geq k^d$ for all \mathbf{m} .

(ii) If $1 - 3q + q^2 < 0$, we still have $k^c = 3 > k^d = 3\mathbf{b}$ for

$1 + \mathbf{d}[1 - (2q - q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}]$. For $1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}] < \mathbf{m} \leq 1 + \mathbf{d}$, it

can be verified that $\frac{3}{q}[1 - (1 - \frac{1 + \mathbf{d} - \mathbf{m}}{\mathbf{d}\mathbf{g}})^{\frac{1}{2}}] = 2$ when $\mathbf{m} = 1 + \mathbf{d}[1 - \frac{4}{9}(3q - q^2)\mathbf{g}]$. If

$-\frac{4}{9}(3q - q^2) \leq 1 - 3q + q^2$, or $q \leq 0.83$, then $k^c \geq k^d$ for all \mathbf{m} . If $-\frac{4}{9}(3q - q^2) > 1 - 3q + q^2$,

then $k^c = 2 < k^d = 3\mathbf{b}$ when $1 + \mathbf{d}[1 + (1 - 3q + q^2)\mathbf{g}] < \mathbf{m} < 1 + \mathbf{d}[1 - \frac{4}{9}(3q - q^2)\mathbf{g}]$. **Q.E.D.**

Proof of Lemma 3: Given other localities' strategies, the payoff to an individual locality of

experimenting is: $V^E = qS - (1 - q)F + \mathbf{d}\left[qS + (1 - q)\left(1 - (1 - \mathbf{b}q)^{(N-1)}\right)\mathbf{g}S\right]$, and the payoff

from not experimenting is: $V^{NE} = \mathbf{d}\left(1 - (1 - \mathbf{b}q)^{(N-1)}\right)\mathbf{g}S$. In the mixed strategy equilibrium, it

must be that $V^E = V^{NE}$ and $0 < \mathbf{b} < 1$. Some straightforward algebra yields the lemma. **Q.E.D.**

Proof of Proposition 4: When $\mathbf{b} \in (0, 1)$, it is easy to see that $\mathbf{b} \rightarrow 0$ as $N \rightarrow \infty$. As $N \rightarrow \infty$, it

can also be verified that $\hat{k}^d = N\mathbf{b} \rightarrow -\frac{\ln(1 - \frac{1 + \mathbf{d} - \mathbf{m}}{\mathbf{d}\mathbf{g}})}{q}$ which is a constant. **Q.E.D.**