Constitutions and Policy Comparisons:*
Direct and Representative Democracy When States Learn From Their Neighbours

David Hugh-Jones

Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ

Email: dhughj@essex.ac.uk

May 20, 2008

Abstract

Voters in democracies can learn from the experience of neighbouring states: about policy in a direct democracy ("policy experimentation"), about the quality of their politicians in a representative democracy ("yardstick competition"). Learning between states creates spillovers from policy choice, and also from constitutional choice. I model these spillovers in a simple principal-agent framework, and show that voter welfare may be maximized by a mixture of representative and direct democratic states. Because of this, empirical work examining voter welfare under direct democracy may need to be reinterpreted. Also, I show that the optimal mix of constitutions cannot always be achieved in a constitutional choice equilibrium involving many states.

Keywords: policy experimentation, yardstick competition, constitutional choice, direct democracy

*. David Hugh-Jones is a final year PhD candidate at the University of Essex. ADDRESS: Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom. His dissertation is on the institutions of direct democracy. He thanks Hugh Ward, Thomas Plümper, Vera Tröger, Katja Mirwaldt, Kaori Shoji, Haifeng Huang, Jan Schnellenbach, participants in the Essex Political Economy seminar and the 2008 annual meeting of the Public Choice Society, Jim Rogers and two anonymous reviewers for their comments. All mistakes are his own.

1 Introduction

On June 6th 1978, voters in California passed Proposition 13, slashing property taxes. Proposition 13 was a citizen's ballot initiative sponsored by the larger-than-life Howard Jarvis (slogan: "I'm mad as hell, and I'm not going to take it any more!") By the general election in November that year there were similar measures on the ballot in 13 US states. Many were "initiative states" which, like California, allow their citizens to propose legislation to the voters directly. But states without the ballot initiative were also affected. Legislators in Texas and Hawaii brought tax-capping constitutional amendments before the voters; Jarvis traveled to Texas to rally support. A special session of the Alabama legislature proposed ceilings on property taxes which were passed by the voters. In Nebraska the legislature enacted a statutory property tax cap. In Illinois, Iowa, Kansas, New Hampshire, Pennsylvania and Wisconsin, the legislature considered but did not pass tax limitations. Several other states took relatively minor actions such as establishing a commission to examine property taxes (Hawkins 1979, ACIR 1979). Whether or not Proposition 13 helped bring Reagan to power, as some argue (Kirlin 1982; see Smith 1998), it certainly affected legislative behaviour in states across the US, even in those which lacked direct democratic institutions.

In 2005, two Californian ballot initiatives sought to reduce the cost of prescription drugs for low-income families. Proposition 79 was supported by citizens' groups and organized labour. It mandated the Department of Health Services to negotiate collective discounts on drug prices with pharmaceutical companies; uncooperative companies could be shut out of the state Medi-Cal program. Proposition 78 was sponsored by the drug companies. It covered fewer families and contained no penalties for non-participators. The competition between these two propositions was intense and acrimonious. Advocates for either side pointed to evidence from similar recent programmes in Maine and Ohio. The Maine legislature had originally developed "Maine Rx" in 2000. This plan, like Proposition 79, allowed non-participating drug companies to be barred from Maine's Medicaid programme. Maine Rx had been struck down by the courts and in 2003 a voluntary, compromise system was in place. In Ohio, consumer advocates had filed a ballot proposal setting up a similar system. Again, following a court battle, a voluntary compromise system had been set up, though at the time of the California campaign relatively few Ohio residents had been enrolled (California Healthcare Foundation 2005). In this case, then, a direct democratic election campaign – indeed the text of Proposition 79 itself – was shaped by the example of Maine's legislatively developed program.

These examples show what this paper theorizes. Citizens can learn from neighbouring states[1] about the effects of policy, and about the quality of their politicians. In the jargon, political decisions have cross-border information externalities. Learning from your neighbours about policy is known in the literature as "policy experimentation", and "yardstick competition" is what results when voters compare their representatives with those in neighbouring states. The relative importance of these two effects depends on a state's political system – its constitution in the broadest sense. If politicians are entrusted with discretion to make policy, then voters will be concerned to judge their motivations and competence, and yardstick competition will predominate. Politicians themselves may learn about policy from their neighbours, but if they have greater expertise in the first place, this will be less important. In other states, where politicians are like delegates, kept on a tight rein by institutions like the ballot initiative or simply by a less trusting citizenry, policy will cleave closely to the voters' opinions. As voters have little rational incentive to delve into policy details, information from neighbouring states will be more valuable, and policy experimentation will be the most important effect. At the extremes of this constitutional continuum lie the ideal types of representative democracy, in which all policy-making power lies with elected officials, and direct democracy, in which voters themselves always decide. The flow of information across borders affects the tradeoffs between these idealized institutions. This paper examines those effects. There are three main results.

First, because different kinds of democracy produce and consume information in different ways, they affect (and are affected by) their neighbours differently. In particular, representative democracies may be more informative to their neighbours than direct democracies, for the following reason. Policy experimentation produces useful information for all states, but it is in citizens' interests to let other states bear the cost of the experiment, which can go wrong. Thus, direct democracies experiment too little: even when voters can compare many direct democracies, the best policy may not become known. By contrast, when elected representatives have the knowledge to implement a good policy, but may not want to, neighbouring representative democracies may discipline them via the mechanism of yardstick competition. Here there is no need to experiment: politicians know that bad policy choices will be punished by the voters, so always choose good ones. So yardstick competition always works to reveal the best policy, which other states can then copy. Thus, in this paper, more information flows from representative democracies than from direct democracies; existing literature, discussed below, focuses mainly on information coming from direct democracies.[2] Now, because representative democracies provide more information, direct democracies may outperform representative democracies, but only in the presence of representative democratic neighbours. As a result, it may be best for voters to have a mixture of direct and representative democratic systems. Much of the current literature compares these two forms of democracy and tries to discover, free of context, which one is better for voters. According to this paper, the answer could well be “it depends”. One form of democracy may currently be better for voters, but the outcome if more states adopted this form could still be worse for everybody.

The second result follows from the first. Because the benefits of direct democracy decrease as the number of direct democracies increases, if states choose their constitutions independently of one another, a mixture of constitutional forms can be reached in equilibrium. This may help to explain why the ballot initiative institution spread across the US, but stopped almost completely after being adopted in 24 states.

Finally, the mix of constitutions that gets chosen in equilibrium is not always welfare-maximizing. The reason is that constitutional choice, just like policy choice, has externalities across states. An optimal outcome can then only be achieved if states cooperate in choosing their constitutions, which is unlikely unless there is a central governing body.

In the next section, I discuss this paper’s contribution to the existing literature. Section 7 sets out the model. Section 7 shows equilibrium policy choices when there are one or more direct democracies, one or more representative democracies, or a mixture of both kinds, and shows how information spillovers affects the trade-off between the constitutions. Section 7 examines constitutional choice when each state takes its neighbours’ constitutions as given, and explores the optimal mix of constitutions for voter welfare. This contains the main results. Section 7 considers relaxing some of the model’s assumptions. The conclusion sums up and suggests directions for further research.

2 Existing literature

The political science literature on direct democratic institutions is organized around the comparison with traditional representative democracy. This comparison can be undertaken from many points of view. Here I focus on voter welfare.

Theoretically, Gerber (1996) shows that in a complete information setting, policies are closer to the median voter when representatives face the threat of a popular initiative. To analyse the potential benefits of representation, incomplete information is needed: in general, representatives should know something voters don't. Maskin and Tirole (2004) use a principal-agent model in the tradition of Barro (1973) and Ferejohn (1986) to analyse the trade-offs between different forms of democracy given different parameters. I use a very similar framework to show how those tradeoffs shift when policy outcomes can be observed in neighbouring states.

The empirical literature on direct democracy has developed alongside and in response to the theory. The standard format is a comparison of different political units within federations, either in the US or Switzerland. Most of these studies provide evidence that US states with the ballot initiative have policies closer to the will of the majority (Gerber 1996, 1999; Arceneaux 2002, Bowler and Donovan 2004, Matsusaka 2004, Burden 2005), although the finding is not universal (Lascher et al. 1996, Camobreco 1998; but see Matsusaka 2001). A related question is whether policy in these states is better in some objectively measurable sense. Again, the answer is positive (Feld and Savioz 1997, Feld and Kirchgässner 2000; Frey and Stutzer 2000 find that citizens are happier in Swiss cantons with the citizen's initiative).

To sum up, political science theory and empirical work mainly support the view that direct democratic institutions improve voter welfare. But this gives rise to an important positive question. As Matsusaka (1999 p. 133) points out: "the models imply that voters are *always* better off... having the initiative process available. If so, then why do only half of the states and cities in the country have it?" [Italics in original.] This paper offers an answer: the existing mix of constitutions affects the comparison. Even if voters are better off in initiative states, an increase in the number of those states might make all voters worse off; in fact, there may be too many initiative states in equilibrium. If so, the empirical work done so far, though interesting and important, does not allow us to draw policy conclusions about whether direct democracy should be extended.

The key processes modelled by this paper are policy experimentation and yardstick competition. There is a literature on each. Theoretically, policy experimentation does not require decentralized policy choice: a social planner could choose different policies for different states. Indeed, Rose-Ackerman (1980) and Strumpf (2002) investigate the incentives for regions to experiment in a decentralized system, and show that there is too little experimentation. In Strumpf the cause is an information externality. That result reoccurs here in a different framework (Section 4.3). The idea of yardstick competition was introduced into political economy by Salmon (1987). Some recent papers have compared its workings under different constitutional forms. Wrede (2001) examines different party systems. Schaltegger and Küttel (2002) argue informally that within a representative framework, direct democratic institutions will reduce yardstick competition by mitigating representative democracy's incentive problems. This paper brings together yardstick competition and policy experimentation in a single framework, and formally shows how they affect the tradeoffs involved in delegating power to representatives.

An empirical research tradition within political science examines policy diffusion between nations or federal political subunits (see e.g. Berry and Berry 1999). Social learning is recognized as a key cause of policy diffusion, but normally in this literature the learners are bureaucrats or politicians, not voters. The idea that the electorate may learn from other states is present in an early paper (Walker 1969), but until recently, the policy diffusion literature has ignored yardstick competition and focused more on policy experimentation as a theoretical model. However, some recent papers mention yardstick competition (Bailey et al. 2004; Rincke 2007). In particular, the diffusion of morality policy is probably better explained by yardstick competition than by policy experimentation. For example, Haider-Markel (2001) describes how in the 1990s, campaigning by conservative religious groups caused anti-gay-marriage legislation to spread extremely fast through US states. I would argue that legislation spread quickly because, once one state passed the law, legislators in other states feared being exposed as "anti-family" by comparison if they did nothing. A formal model integrating both policy learning and yardstick competition could provide a basis for empirical work on this process.

Matsusaka (2004) discusses the possibility that well-intentioned representative politicians learn about voter preferences from direct democratic institutions. Boehmke (1999) models this process. Politicians learn from their neighbours about voter support for a potential initiative in their own state: in an empirical examination of casino gaming, he shows policy diffusion occurring only between initiative states. In the present paper, there is policy diffusion from representative to direct democracies; direct democracies also have the potential to discipline their representative neighbours, as politicians fear being removed from office if they do not copy popular policies.

Finally the paper adds to a growing theoretical literature that endogenizes the process of constitutional choice: see for example Aghion et al. (2002), Barbera and Jackson (2004). An innovation here is to focus on the international context of constitutional choice. This seems likely to be an important determinant, especially in small and globalized countries.

3 The model

The model uses a simple principal-agent framework. Policy is a choice between two distinct alternatives. Voters share common interests, but do not know which policy will best serve those interests. Politicians know the best policy for the voters, but may themselves prefer the other policy. There are one or more states. Voters observe policy choices, and their results, in other states as well as their own, but the effects of policy are observed imperfectly: voters cannot entirely distinguish bad policy from bad luck. However, "luck" – a random shock to voter welfare – is correlated perfectly between states, so that when states choose different policies, the difference in outcomes reveals the best policy. There are thus two commonalities between states: the best policy and the shock. These strong assumptions help to bring out the intuition of the model: Section 7 considers what happens when they are relaxed.

Formally, states are numbered 1, ..., n, with typical member j, and relevant variables (u_1, u_2, d_1, d_2, r, δ) are indexed with superscripts, which are dropped when the sense is clear. There are two periods. In each period $t = 1, 2$, a policy d_t^j is chosen from two possible options, a or b. One of these two policies is better than the other: call the best policy x. x does not change between periods. With probability $\gamma \geqslant 1/2$, $x = a$, otherwise $x = b$. A high value of γ represents an "easy" policy decision in which common sense or well-known facts favour one policy over the other. The electorate does not observe x.

After period 1 the electorate gets utility from the policy and from a random shock: $u_1 = \delta + \varepsilon$, where $\delta = 1$ if $d_1 = x$, $\delta = 0$ otherwise, and ε is a mean zero random variable with cdf $\Phi(\,\cdot\,)$ and pdf $\varphi(\,\cdot\,)$. We make some technical assumptions about the shock: φ is symmetric and differentiable, has full support on $\mathbb{R}$, and is single-peaked at 0, and the induced distribution of u_1, with cdf $\Theta(u) = \Phi(u - \delta)$ and pdf $\theta(\,\cdot\,)$, satisfies the monotone likelihood ratio property with respect to δ. These will all be satisfied if, for example, ε is normally distributed. In period 2, the electorate receives utility $u_2 = 1$ if $d_2 = x$, $u_2 = 0$ otherwise: we ignore any period 2 shock to utility. The assumption of common interests can be interpreted as follows: a majority of voters prefer x, and δ represents the difference in total utility between serving the majority and serving the minority.

The electorate in each state observes period 1 policy decisions, and the resulting voter utility levels, in all states. This learning process need not demand unrealistic levels of citizen political engagement; instead, it can be thought of as a reflection of a well-functioning democratic system, in which relevant information is publicized by interest groups, political parties, the media, and so on. The number of states n is a simple way to represent the effect of having more states to compare outcomes over: more complex structures would be possible, for example with states only learning from geographic neighbours. The utility shock ε and the best policy x are common between states. Therefore, when policies differ, *all electorates will learn which is the better policy* by observing higher utility in some states. If policies are the same, on the other hand, no extra information is learned. Having x common between states is one end of a possible continuum of assumptions: at the other end, the best policy could be completely independent between states, in which case choices in neighbouring states would bear no lessons for voters (though they might be able to infer the value of the shock ε from observing utility). The assumption of a common shock ε can be thought of as reflecting the benefits from having "variation in the independent variable", when potential policy utility is in fact unknown. That is, citizens possess some information about the utility of the status quo, simply because they experience it in their daily lives. But they lack the opportunity or motivation to learn what their utility would counterfactually be under the alternative policy. If the alternative is tried elsewhere, this information becomes freely available, and can be reported via the media, for example. To be clear, I am not assuming that all citizens pay careful attention to the policies and experience of different states. The Condorcet Jury Theorem shows that even low levels of individual citizen information can result in almost certainly correct decisions after aggregation by a democratic vote. Rational ignorance is therefore a problem of information *production*, not of information *aggregation*. If the experience of neighbouring states provides costless information about a policy, this can alter the voting outcome even if only a few voters are aware of that information. Nevertheless, having a common shock and common best policy are strong assumptions, and I discuss the effect of loosening them in Section 7.

I examine two kinds of constitution. In a direct democracy (DD), the electorate chooses d_1 and d_2 for itself. In a representative democracy (RD), a representative decides policy. The representative knows the correct policy x with certainty. This greater knowledge reflects the fact that representatives have access to a bureaucratic staff, and choose policy unilaterally, whereas individual voters are rationally ignorant because gathering information is costly and individual votes are unlikely to affect the outcome. The representative has the same preferences as the electorate over policy , i.e. prefers $d_t = x$, with probability π. Such a representative is called "congruent" or simply "good". Otherwise the representative prefers $d_t \neq x$ and is "non-congruent" or "bad". Write $r^j \in \{a, b\}$ for the jth representative's preferred alternative. If $r^j = x = k \in \{a, b\}$, we describe the representative as a *congruent k-type*; if $r^j \neq x = k \in \{a, b\}$, we call her a *non-congruent k-type*.

At the broadest level, the possibility of non-congruent politicians is meant to reflect the idea that elected officials do not always represent their citizens' interests perfectly. A more specific interpretation, for $\pi \geqslant 1/2$, would be that a proportion π of citizens prefers option x, while the minority prefers the other option: the utility difference from choosing the majority-preferred option is normalized to 1, and the politician is drawn at random from the citizens. In this model the probability of a politician being congruent is exogenous. It would be more realistic (but also more complex) to assume that politicians only run for office if this offers higher expected utility than some alternative career. If so, when voters were better able to distinguish good from bad politicians, better politicians would run for office. This would add to the positive effect of yardstick competition, by selecting bad politicians out before the political process started.

At the end of period 1 the electorate chooses either to eject the incumbent representative and choose a new one, who will again be congruent with probability π, or to keep the incumbent. Representatives get utility from staying in office, and utility from implementing their preferred policy – but only if they themselves implement it while in office. This could reflect either the motivation to leave a political "legacy", or policies that benefit the current office holder only, such as an increase in the representative's salary. (A more standard "policy motivation" assumption, by which representatives care about policy outcomes whoever implemented them, would not change the results.) Let G be the representative's utility from choosing her preferred action, R be the perks of office and $0 < \beta < 1$ her discount rate between periods. Write

$$\kappa = \frac{G}{\beta(R+G)} \tag{1}$$

for the relative benefit of choosing one's preferred policy in period 1 compared to choosing one's preferred policy, and getting the perks of office for one more period, in period 2. Two common motivational assumptions can be accommodated within this framework: if $G = 0$ so that representatives are purely office-motivated, then $\kappa = 0$; if $R = 0$, representatives care only about their "legacy" and $\kappa = 1/\beta > 1$. If $\kappa > 1$, I say that the representative is *impatient*, because she is more concerned about the immediate benefit of her preferred policy than about the delayed benefits from office in period 2. In order to restrict off-equilibrium beliefs and reduce the number of equilibria, I sometimes assume that a small proportion of representatives is impatient, whatever the value of κ in general.

If a representative chooses $d_1 \neq x$, I say that the representative *shirks*, and if a non-congruent representative chooses $d_1 = x \neq r$, I say that the representative is *disciplined*. Only non-congruent representatives shirk, but the converse is not true.

In this finitely repeated framework, representatives always choose their own preferred policy in period 2. This motivates voters' desire to reelect only congruent representatives, whatever the equilibrium. Substantive results would be similar in a repeated game where representatives always faced a further election: under realistic assumptions, there are still going to be occasions (e.g. close to retirement) when representatives will follow their own preferences rather than those of the voters, and if so then congruent representatives will always be preferable.

Some general remarks will help explain the model. The first-best outcome is clearly for x to be chosen in both periods, giving utility of 2 if we assume no discounting between periods, which is a reasonable approach for normative analysis in this case.[3]

In a direct democracy, the electorate's best period 1 choice is to follow its prior and choose $d_1 = a$, giving period 1 utility of γ. After period 1, some information is learned. At worst, if no learning takes place, the electorate can choose $d_2 = a$ and achieve expected utility of 2γ. At best, we could hope for the right policy x to be discovered with certainty. Without discounting between periods, expected utility would then be

$$\gamma + 1. \tag{2}$$

This provides an upper bound on performance from a direct democracy.

In a representative democracy, representatives always follow their preference in period 2, $d_2 = r$, as they face no further elections. One possible outcome is that in period 1, representatives always choose $d_1 = x$, perhaps because of the threat of losing the election if they do not do so. Voter welfare is then

$$1 + \pi\,; \tag{3}$$

as all representatives behave equally well in period 1, there is no way for voters to distinguish them and eject non-congruent representatives. On the other hand, maybe all representatives choose $d_1 = r$. In particular if $\kappa > 1$ then representatives will certainly choose $d_1 = r$, as $\kappa > 1 \Leftrightarrow G > \beta(R+G)$ means that the immediate benefit of following one's preference outweighs even the certainty of electoral defeat. Then the electorate faces a problem of detecting and ejecting non-congruent representatives, and retaining congruent ones. Say that a non-congruent representative is detected and ejected with probability X. It may also be that a congruent representative is falsely thought to be non-congruent and ejected. Let Y be the probability of this kind of "false positive". Voter utility will then be

$$\begin{aligned} &\pi[1+(1-Y)+Y\pi] \\ &\quad +(1-\pi)[0+X\pi] \\ &=2\pi+\pi(1-\pi)(X-Y), \end{aligned} \tag{4}$$

recalling that a new representative has probability π of being congruent.

Expressions (3) and (4) reveal a tradeoff between moral hazard and adverse selection. When representatives are undisciplined in period 1, the voters suffer from moral hazard: their agent, the representative, may take the worse action for them. However, the difference between good and bad representatives' actions allows them to detect at least some bad representatives. On the other hand, if representatives are disciplined in period 1, voters suffer from adverse selection: they cannot weed out bad representatives who are then free to follow their preferences in period 2. Because replacement representatives are themselves not always congruent, the gain from detecting bad representatives who shirk is never enough to compensate for the period 1 loss. Even in the best possible case, $X-Y=1$, (4) evaluates to $2\pi + \pi(1-\pi) < 2\pi + (1-\pi) = 1+\pi$. So, in general, (3) provides an upper bound on the performance of representative democracy, and (4) provides an upper bound on RDs when representatives shirk.

The analysis focuses on the best available equilibrium for the voters, in order to examine the different institutions at their peak performance. Representatives in different states could coordinate on an equilibrium that is best for them not the voters, but a full exploration of this requires a paper in its own right and is left for future work (see the conclusion). There is one exception to this rule: when there are many direct democracies (Section 4.3), I examine a mixed equilibrium rather than a more efficient, but implausible, asymmetric pure strategy equilibrium.

4 Policy choice

I first analyse equilibrium when $n = 1$, as a benchmark, then look in turn at the cases of more than one direct democracy, more than one representative democracy, and a mixture of systems.

4.1 A single direct democracy

Our first proposition describes the unique equilibrium under direct democracy when there is only one state. In this setting, the electorate chooses $d_1 = a$, in accordance with its prior belief, and then stays with option a unless its period 1 utility is below a particular cutpoint $\bar{u}$. So with only one state, direct democracy is inefficient because the electorate receives only a noisy signal from choosing policy.

Proposition 1. *When there is a single direct democracy,* $d_1 = a$ *and* $d_2 = a$ *if* $u_1 > \bar{u}$, $d_2 = b$ *otherwise, where* $\bar{u}$ *uniquely solves* $\varphi(\bar{u})/\varphi(\bar{u} - 1) = \gamma/(1 - \gamma)$. *Expected voter utility is* $\gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u})$.

Proof. See the appendix. □

4.2 A single representative democracy

The fundamental issue in representative democracy is whether non-congruent representatives are disciplined by the threat of losing an election, or shirk because they are impatient or because the electorate cannot distinguish congruent from non-congruent representatives accurately enough. In a single state, there is also a third case involving only partial discipline.

Definition. *A* disciplined equilibrium *is one in which all types of representative play* $d_1 = x$ *and* $d_2 = r$. *In a* moral hazard equilibrium, *all types of representatives choose their own preferred action at all times:* $d_1 = d_2 = r$.

Proposition 2. *When there is a single representative democracy:*

1. *There is a disciplined equilibrium if and only if* $\kappa \leqslant \Phi(\bar{u}) - \Phi(-\bar{u})$. *Voter utility in this equilibrium is* $1 + \pi$.

2. *There is a moral hazard equilibrium if and only if* $\kappa \geqslant \Phi(1 - \bar{u}) - \Phi(\bar{u} - 1)$. *Voter utility is* $2\pi + \pi(1 - \pi)\{\Phi(\bar{u}) - \Phi(\bar{u} - 1)\}$.

3. *If and only if* $\Phi(\bar{u}) - \Phi(-\bar{u}) < \kappa < \Phi(1 - \bar{u}) - \Phi(\bar{u} - 1)$, *there is an equilibrium in mixed strategies: congruent types play* $d_1 = x$, *while non-congruent types play* $d_1 = a$ *(shirk) when* $x = b$ *and mix between* a *and* b *when* $x = a$.

Proof. See the appendix. □

The condition for a disciplined equilibrium depends on three factors: the level of representatives' impatience κ, the form of the utility shock distribution Φ, and the ease of the issue γ (which determines $\bar{u}$ given φ). Clearly a higher level of impatience makes a disciplined equilibrium harder to achieve. As $\gamma \rightarrow 1$, $\bar{u} \rightarrow -\infty$ and again it becomes harder to achieve a disciplined equilibrium. In particular, if $\bar{u} < 0$ it will be impossible to fulfil condition 1. The intuition here is that when γ is high, it will take a very low utility level to persuade voters that the best policy is b. So a non-congruent b-type will face a big temptation to defect. Finally, when the variance of the utility shock is high, discipline will be harder to achieve as it is harder for voters to know when they are being cheated.

4.3 Many direct democracies

Suppose now that there are $n > 1$ states, all direct democracies. If at least one state chooses each policy, all electorates can learn the best policy with certainty after the first period. But an electorate that chooses the less likely policy b is going against its prior. If a is very likely to be the best policy, the electorate will prefer to choose $d_1 = a$ and bear the resulting loss of information. This information is also lost to all other states, and as a result the equilibrium is inefficient. Even if the prior γ is low enough that a single electorate in state j, faced with all other states choosing a, would prefer to choose $d_1^j = b$, one would not necessarily expect this to happen, because all electorates will prefer that some other state pay the cost of choosing b. In fact, there is a mixed equilibrium in which all states randomize between policies. As a result, sometimes all states choose the same policy, normally a, and are unable to learn the best policy with certainty. Thus, direct democracies are unable to fully exploit the advantages of cross-border information for policy experimentation, because no single electorate wants to experiment on behalf of all the others.

Proposition 3. *When there are $n > 1$ direct democratic states,*

1. *if and only if $\gamma < \bar{\gamma}$ where $\bar{\gamma}$ uniquely solves $\bar{\gamma} = \frac{1 + \Phi(-\bar{u})}{2 - \Phi(\bar{u} - 1) + \Phi(-\bar{u})}$ there is a symmetric mixed equilibrium in which all electorates choose $d_1 = a$ with probability $\alpha \in [1/2, 1)$; for any fixed $\gamma < \bar{\gamma}$, α increases towards 1 as $n \to \infty$, and $\alpha^n \to \frac{2\gamma - 1}{\gamma\Phi(\bar{u} - 1) + (1 - \gamma)\Phi(-\bar{u})} > 0$ as $n \to \infty$.*

2. *otherwise, all states play* $d_1 = a$ *with certainty, giving expected voter utility of* $\gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u}) < \gamma + 1$.

Proof. See the appendix.[4] □

Here α^n gives the probability that all electorates choose a in the first period. When this happens, learning is not complete and the wrong policy may be chosen in the second period. As α^n is positive, no matter how many states there are, there is always a positive probability of this miscoordination, so that direct democracies on their own never achieve the best possible voter utility of $\gamma + 1$.

This inefficiency naturally brings up the question of side-payments. If there were a mechanism which let some states pay others for making a risky policy experiment, then the electorates could agree a suitable payment and ensure that some state chose b in period 1. Doing this directly via agreements among electorates – perhaps a series of popular votes mandating a payment of X if all other states offer the same – seems unrealistic given the transaction costs involved. A benevolent central government could coordinate payments via taxes, but there may not be a central government for the group of states concerned. Also, the model assumes that politicians are self-interested. Extending this assumption to central government, centralization may be more about allowing collusion between politicians than about solving voters' coordination problems.

4.4 Many representative democracies

Representatives are still congruent with probability π, and we assume that congruence is independent between states.[5] If all states use representative democracy, then as before there are equilibria in which representatives are disciplined to choose $d_1^j = x$, and moral hazard equilibria in which they follow their preferences. However, the condition for discipline is weaker, because if the representative in any other state is disciplined and choosing $d_1 = x$, then a choice of $d_1 \neq x$ will be revealed with certainty by the utility difference. So a non-congruent representative stands no chance of reelection if she chooses to follow her own preference. Also, even if representatives are impatient enough for a moral hazard equilibrium, the possibility of different representatives choosing differently approaches 1 as n gets large. Again, this difference between states will reveal x, so that non-congruent representatives can be detected and weeded out.

Proposition 4. *When there are $n > 1$ representative democracies, iff $\kappa \leqslant 1$, there is a disciplined equilibrium in which voter welfare is $1 + \pi$. If $\kappa \geqslant K(n)$, where $K(n) < 1$ and $K(n) \to 1$ as $n \to \infty$, then there is a moral hazard equilibrium and voter welfare approaches $2\pi + \pi(1 - \pi)$ as $n \to \infty$.*

Proof. If representatives in all states j choose $d_1^j = x$, a non-congruent representative who unilaterally deviates is certainly detected because the electorate observes the difference in utility levels between the states. Specifically, if there is a small proportion of impatient representatives who always follow their own preferences and choose $d_1^j = r$, then given equilibrium strategies, the electorate in state j will conclude, on observing $d_1^j \neq d_1^k$ and $u_1^j < u_1^k$ for some state k, that their representative is impatient and non-congruent, and eject her. Thus, the condition for this equilibrium is simply $G \leqslant \beta(R + G) \Leftrightarrow \kappa \leqslant 1$. As representatives are disciplined, voter welfare is $1 + \pi$ as in (3).

Suppose that all representatives choose $d_1^j = r^j$ where r^j is the state j representative's own preference. Then, if $d_1^j \neq d_1^k$ for some $j, k \in \{1, ...n\}$, non-congruent representatives will be revealed by the utility difference, and they will be ejected with certainty (and congruent representatives will be certainly reelected). The probability of this event is

$$1 - \pi^n - (1 - \pi)^n \to 1 \text{ as } n \to \infty. \tag{5}$$

Thus as n grows large the probability of survival in office for a non-congruent representative choosing $d_1^j = r_j$ approaches 0 and the condition to choose $d_1^j = r_j$ becomes $G \geqslant \beta(R + G) \Leftrightarrow \kappa \geqslant 1$. As congruence and non-congruence are almost always detected, voter welfare approaches the best possible for a moral hazard equilibrium: $2\pi + \pi(1 - \pi)$, as in (4). □

Cross-border learning helps representative democracy just as it does direct democracy: voters can observe and make inferences from the difference in outcomes between states. But in a RD the inefficiency of experimentation can be avoided, because the threat alone of discovery motivates representatives to choose the best outcome. In technical terms, the worse policy is chosen only "off the equilibrium path".

Note that while the disciplined equilibrium does not require a large number of states – the conditions will suffice even when there are just 2 – the moral hazard equilibrium approaches maximum welfare only when there are many states. In a system with just a few states, shirking will sometimes go undetected because all representatives chose the same policy, and so a moral hazard equilibrium will exist for values of κ less than one.[6] Thus there are multiple possible equilibria. I assume that when $\kappa \leqslant 1$ the disciplined equilibrium, which gives the highest voter welfare, is selected.

4.5 A mixed system

Finally, I examine the case where states 1, ..., m use representative democracy, while the remaining $n - m$ states use direct democracy. I assume that there are least two representative democracies. As before, if $\kappa \leqslant 1$, there is a disciplined equilibrium in which all representatives choose $d_1^j = x$ for $j \in \{1, ..., m\}$. The condition for this is simply that $m > 1$; as shirking is always detected anyway, the presence of DD states makes detection no easier. But the RD states do help voters in the DDs, who can infer x with certainty from representatives' period 1 choices and equilibrium strategies, and can therefore follow their prior and choose $d_1 = a$, achieving their best possible utility of $\gamma + 1$.

This point is key. Section 4.3 showed that, in the presence of other direct democracies, a direct democracy will not achieve its maximum voter utility, because of the costs of experimentation. But if there are representative democracies, and their representatives are disciplined to choose the right policy, then the direct democracy can copy from them and achieve maximum voter utility. In other words, direct democracy is better for voters when there are representative democracies to learn from, than when there are only other direct democracies.

In a moral hazard equilibrium when $\kappa > 1$, the presence of direct democratic states again makes little difference to RDs as impatient representatives prefer the benefits of shirking even if they are detected with certainty. If m is sufficiently large, the direct democracies will again be able to infer the correct policy – this time by observing the utility difference between representative states, some of whose representatives will almost certainly choose either policy option. Thus again they can follow their prior in period 1 and achieve expected welfare of $\gamma + 1$. So, the presence of representative democracies allows direct democracies to avoid the inefficiency loss from not experimenting. The following proposition sums up our discussion.

Proposition 5. *In a mixed system with $m > 1$ representative democratic states, there is an equilibrium iff $\kappa \leqslant 1$ in which all representatives are disciplined, $d_1^j = x$ for $j \in \{1, ..., m\}$ and direct democratic states choose $d_1^j = a$ for $j \in \{m+1, ..., n\}$. There is an equilibrium in which non-congruent representatives shirk, $d_1^j = r^j$ for $j \in \{1, ..., m\}$, if $\kappa \geqslant 1$; in this equilibrium, for large enough $n - m$, again $d_1^j = a$ for $j \in \{m+1, ..., n\}$.*

Average voter utility in a mixed system equals

$$(1+\pi)\frac{m}{n}+(1+\gamma)\frac{n-m}{n} \tag{6}$$

when $n-m>1$ and $\kappa\leqslant 1$, and approaches

$$(2\pi+\pi(1-\pi))\frac{m}{n}+(1+\gamma)\frac{n-m}{n} \tag{7}$$

when $\kappa>1$ and m grows large. In these cases, direct and representative democracies are performing at or near their best, giving the highest possible utility levels. Because there are many of both kinds of democratic institution, every state benefits as much as possible by learning from both kinds.

5 Constitutional choice and voter welfare

Constitutions are not fixed. Ideally, a state's constitution is chosen by its citizens. But they make this choice in a context partly determined by other states' actions. Democratization proceeds in waves, partly because the democratic forces in any country are likely to be encouraged by the success of nearby revolutions (Huntington 1991; Boix 2003). Similarly, the populist and progressive movements introduced the initiative in 19 US states during the first quarter of the 20th century (Cronin 1989). Here, I make the very simple assumption that voters are able to choose the constitution they prefer. The context I focus on is provided by the information available from other states during the normal policy-making process. States with different democratic institutions learn differently from their neighbours, and the learning process also depends on their neighbours' constitutions. In particular, as the last section showed, a direct democracy gains from having many representative democratic neighbours. So, when all other states are RDs, direct democracy may be preferable to representative democracy, but when all other states are DDs, the reverse may be true. If so, there will be an equilibrium involving a mixture of costitutions.

To clarify this formally, I define a simple notion of stability for a given set of constitutions.

Definition. *A system of direct and/or representative democracies is* stable *if each state's electorate prefers the state to keep its current constitution, given that all other states do the same. Preference is defined in terms of expected utility from the best possible equilibrium in the policy choice game.*

We can interpret this idea of stability in two ways. One interpretation would be that all voters simultaneously choose their states' constitutions. Stability then just means that there is a Nash equilibrium in their choices of constitution. Another interpretation is closer to historical reality. Suppose that, at some time, voters in a single state choose which form of democracy to use, perhaps by holding a relatively rare constitutional convention, and that they ignore any future changes in other states' constitutions, perhaps because voters have relatively short time horizons. For example, this corresponds roughly to how US states have revised their constitutions over time. In this case, a stable system is one in which no state will change its constitution on its own.

The next Proposition shows that for certain parameter values, only a mixed system can be stable. For simplicity's sake I focus on the case when γ is large enough that all direct democracies choose policy a in period 1, and on extreme values of κ, although the logic would continue to hold for lower values of γ and intermediate values of κ.

Proposition 6. *1. For $\kappa < \Phi(\bar{u})$ and any $\gamma > \bar{\gamma}$, there is some π such that neither a system of representative democracies nor a system of direct democracies is stable.*

2. For $\kappa > 1$ and any $\gamma > \bar{\gamma}$, there is some π such that when n is large enough, neither a system of n representative democracies nor a system of n direct democracies is stable.

Proof. See the appendix. □

I next analyse voter welfare. The intuition is as follows. If representative democracy when politicians are disciplined outperforms direct democracy at its best, then to maximize voter welfare we ought to have all representative democracies, for then politicians will be disciplined very easily by their neighbours, and this will be better in every state than having even the best direct democracy. On the other hand, if direct democracy at its best outperforms representative democracy, then we may still want to have a few representative democracies around for the direct democracies to learn from. Otherwise, the direct democracies may not achieve their best, because they are unable to solve the collective action problem of policy experimentation. In general, the positive externality from RDs should decrease with the number of RDs: as correct policy becomes clearer, it is less necessary to have an extra RD to ensure it.

Proposition 7. *When $\kappa \leqslant 1$, and n is high enough, the welfare-maximizing set of constitutions always includes at least one representative democracy, and is composed only of representative democracies if $\pi > \gamma$.*

Proof. Suppose that $\kappa \leqslant 1$. From (6), clearly if $\pi > \gamma$ it maximizes voter welfare to have all RDs, as these will be disciplined and achieve welfare of $1 + \pi$ which is greater than the highest possible welfare in a direct democracy.

If $\pi < \gamma$, and n is large, then when there are two RDs from (6) average welfare is $\frac{2}{n}(1 + \pi) + \frac{n-2}{n}(1 + \gamma)$ which approaches $1 + \gamma$ as n grows large. (If κ is low enough that a single RD would be disciplined, slightly higher welfare can be achieved by having only one RD). On the other hand, a constitution of only DDs will not achieve their maximum welfare of $1 + \gamma$, as they will not learn the right policy with certainty in period 2 (see Proposition 3). Thus for n large enough, it will be better to have at least one representative democracy. □

We might also ask whether the optimal profile of constitutions will be reached in equilibrium. When the optimal profile is a mixture of DDs and RDs, this does not always happen. The reason is analogous to the case of policy experimentation with many direct democracies: no individual state wants to bear the loss of being a RD and providing information to its neighbours.

Proposition 8. *When $\kappa \leqslant 1$, $n \geqslant 3$ and $\pi > \gamma$ then the welfare-maximizing set of constitutions is stable. When $\kappa \leqslant 1$, $\pi < \gamma$ and n is large, the welfare-maximizing set of constitutions may not be stable.*

Proof. If $\pi > \gamma$ and $n \geqslant 3$ then trivially no RD electorate would wish to become a DD and gain utility of $1+\gamma$ rather than $1+\pi$, so a set of all RDs is stable.

If $\pi < \gamma$ and n is high enough, then the welfare-maximizing set of constitutions contains either two or one representative democracies, with welfare of $1 + \gamma$ for DDs and $1 + \pi$ for RDs. Clearly no DD will switch. An example shows that in some, but not all cases an RD will switch. Suppose that $\kappa = 0$ so that the representatives in a single RD are disciplined: thus there is just one RD in the optimal system, by the previous Proposition. Let $\gamma \geqslant \bar{\gamma}$ so that every state will choose $d_1 = a$ in an an all-DD system. Switching to direct democracy will then give our RD utility of $\gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u}) < 1 + \gamma$ (Proposition 3). Thus, there are two cases: if

$$1 + \pi < \gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u})$$

then the RD will switch to DD; if

$$\gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u}) < 1 + \pi < 1 + \gamma$$

it will not.[7] □

What if representatives are not disciplined? As before, there are two possibilities. First, even undisciplined representatives may outperform direct democracy at its best, and if so it will be best to have only RDs. But if direct democracy can outperform undisciplined representatives, it may still be useful to have some RDs because of the chance that they will implement different preferred policies and thus, in effect, do the policy experimentation for the DDs. (This argument depends crucially on representatives' policy choices being varied. If undisciplined representatives all chose the same policy option – perhaps a lazy default – then there would be no gains from experimentation.) However, this means that there is a positive externality from having RDs and in equilibrium there are then normally too few RDs.

Proposition 9. *If $\kappa > 1$ and n is large, and $\gamma > \bar{\gamma}$ as defined in Proposition 3; then*

1. *when $2\pi + \pi(1-\pi) > 1 + \gamma$ then a system of all representative democracies is stable and maximizes voter welfare;*
2. *when $2\pi + \pi(1-\pi) < 1 + \gamma$ then in general the equilibrium number of representative democracies is smaller than the welfare-maximizing number of representative democracies.*

Proof. See the appendix. □

A typical illustration is given in Figure 1. Here the x axis represents the number of representative democracies, and the y axis gives utility. (Parameters were set to $\pi = 0.63$, $\gamma = 0.5$, $n = 10$ and Φ normally distributed with $\sigma^2 = 1$.) The solid and dashed lines give voter utility in RDs and DDs respectively. The equilibrium outcome will be close to where these lines cross: otherwise voters in one kind of state could gain by switching to the other. In this case there are 4 representative democracies in equilibrium, as when there are 3 RDs a DD can gain by changing to an RD. Average voter utility is shown by the dotted line and is maximized with 6 representative democracies.

[FIGURE 1 HERE]

Figure 1. Optimum and equilibrium number of representative democracies

Figure 1 illustrates a point mentioned earlier. In equilibrium, with 4 RDs, voter welfare is higher in the DDs than in RDs. Empirical work in this situation would “demonstrate” the superiority of direct democracy. But voter welfare in all states would *decrease* if there were more direct democracies, and average voter welfare would increase if there were more RDs!

6 Loosening the assumptions

I now consider loosening some of the model's simple assumptions.

Perfect observability of neighbouring states, and the common utility shock, could be weakened to imperfect observability or imperfectly correlated shocks in states. In either case, when policy varied between states, citizens would only gain probabilistic knowledge of x. Perfect discipline would thus no longer require only two states and $\kappa \leqslant 1$; instead, discipline would be easier to achieve (i.e. be in equilibrium for higher values of κ) as the number of representative democracies increased, as each extra RD would make knowledge of x, and thus of representatives' congruence, more accurate. If discipline were achieved, direct democracies would still be able to freeride off their neighbours by observing their policy choices. So our fundamental result of the externality between types of constitutions would remain.

What if the best policy varies between states? After all, without this possibility, decentralized policy-making loses one of its basic justifications (Oates 1999; but for a different line of reasoning, see Besley and Coate 2003). If the best policy were completely independent between states, then observation of neighbours would only give more information about the value of the shock: with a common shock, this might still tell citizens whether the best policy had been chosen, but having policy variation would no longer be necessary or sufficient for this. Disciplined equilibria would still exist, as a deviator would still be revealed by the utility difference between states; direct democracies would thus still benefit from RD neighbours, but a system of DDs would approach maximum utility of $\gamma + 1$ as n grew large, because even if all states choose a, there are likely to be utility differences between states which reveal whether each has chosen its preferred option. Thus, for all-DD systems to be suboptimal, there must be at least some correlation between the best policy in each state; the higher this correlation, the more slowly maximum utility is approached as n increases, because learning from utility differences is less likely when all states choose a. Finally, if both x and ε are independent between states, then of course citizens simply learn nothing from their neighbours.

A key assumption when representatives are undisciplined is that they may shirk to either option. An alternative approach would be to assume that shirkers always choose one option, say a. For example, arguably it is always easier for representatives to increase borrowing rather than taxation, as borrowing affects future citizens. Under this approach, shirkers will not be detected when the best option is indeed a, no matter how many states there are. So the minimum κ to allow a moral hazard equilibrium will not increase towards 1 as n grows large. Nevertheless, the presence of RDs continues to benefit DDs when n is large: DDs will either observe all representatives choosing a, and conclude that almost certainly $x = a$, or will observe variation with congruent representatives choosing b, and learn that $x = b$. Exploring this option leads naturally to the possibility of representatives colluding, perhaps via side payments, to enforce a uniform policy. This is discussed further below.

In general, then, the externalities between states are quite robust to loosening the model's stylized assumptions. A more fundamental change would be to have conflict of interest between citizens. If the policy choice remains binary, then so long as the majority-preferred policy also maximizes social welfare, conflict can be accommodated within the model by assuming that 1 represents the utility difference between satisfying the majority and the minority. But it would also be useful to model a richer set of policy alternatives, with a continuous policy space, and a common shock to the median voter's optimal policy, observed only by politicians. Intuitively, yardstick competition would still discipline representatives towards the median voter, while direct democracies would face the same externalities as before in experimenting with different positions.

7 Conclusion

All states must choose what level of control to exert over elected representatives – to subject them tightly to the popular will, and risk losing their expertise, or to give them free rein and risk the effects of their self-interest. This choice is made partly at the constitutional level, and the best choice may depend on the level of information provided by the environment, in particular by other states' choice of policies. Learning from other states creates two benefits. First, citizens themselves can observe the effects of policy and may be in a better position to make decisions themselves: this is the benefit of policy experimentation. Second, citizens can judge their elected officials against those in neighbouring states. This is the benefit of yardstick competition. But there is an asymmetry between these benefits. Citizens' rational ignorance of the best policy can only be removed by actual experimentation. And because one state must bear the risk of an untried policy, there may be a collective action problem which leaves the gains from experimentation unrealised. On the other hand, self-interested representatives may be deterred from bad choices by the threat of being caught out, so that the gains from avoiding moral hazard can be had without any state losing out. This is particularly likely if there are many representatives, all of whom discipline each other and prevent any of their number from stepping out of line. When this happens, voters themselves will be able to trust the policies chosen by representatives in neighbouring states, and may wish to implement them themselves if their representatives are not willing, as when California's Proposition 79 was based on the plan of the Maine legislature. Direct democracy may then outperform representative democracy only if there are enough representative democratic neighbours.

In this context it is suggestive to consider the history of the ballot initiative in the US. The initiative was introduced at a time of deep dissatisfaction with elected representatives, who were seen as corrupt representatives of party politics (Cronin 1989) and who were out of touch with the concerns of voters in the new cities of the mid-West (Matsusaka 2004). 19 states introduced the initiative in the first quarter of the 20th century. But the rush to direct democracy ended as swiftly as it started, and since 1920 only four new states have introduced the ballot initiative. Cronin claims (p. 59) that "even though its most fervent champions often intended less to strengthen representative democracy than to bypass or punish it, it simultaneously helped remedy the defects of representative political institutions." Undoubtedly the major effect the initiative had was on representatives in states which adopted the institution. But if other states were unaffected, it is hard to see why their legislatures were not also forced to accept the initiative. One possible answer is that competition from initiative states – and perhaps the threat of an active and successful direct democracy movement – drove up legislative standards and reduced the pressure for change.

My final result is that because representatives help to discipline each other, there is an externality between states choosing how much power to delegate to them. It may be that electorates in individual states prefer to retain more power than is ideal for them collectively. All states would then benefit from a collective agreement to move towards representative democracy.

The model here probably underestimates the benefit representative democracies receive from yardstick competition, in two senses. First, by providing more information for citizens, neighbouring states make life easier for "good" politicians and harder for "bad" ones. This will affect the incentives to run for office and change the mix of politicians who do so. In the current model the mix of good and bad types is fixed.

Second, if their time horizons are long enough, representatives have opportunities and incentives to collude in order to prevent yardstick competition, just as other oligopolists do. For example, representatives might be able to coordinate on choosing their own preferred policies, rather than those desired by the electorate. In the context of long-term political relationships or institutions like political parties, they may even be able to constrain each other to choose a uniform policy so that no individual politician stands out as especially good or bad. As the story of California's Proposition 13 suggests, direct democracy can break this collusion by demonstrating that certain policy alternatives are workable. This kind of benefit can only be provided by direct democracy, and it might be observed even when direct democracy is not in itself a very effective decision-making mechanism. Direct democracy's positive externalities in preventing collusive behaviour between representatives offer a promising avenue for further research.

Appendix: proofs of propositions

Proposition 1. *When there is a single direct democracy, $d_1 = a$ and $d_2 = a$ if $u_1 > \bar{u}$, $d_2 = b$ otherwise, where $\bar{u}$ uniquely solves $\varphi(\bar{u})/\varphi(\bar{u} - 1) = \gamma/(1 - \gamma)$. Expected voter utility is $\gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u})$.*

Proof. The electorate chooses $d_1 = a$, in accordance with its prior. The electorate's belief that $x = a$ after receiving random utility of u_1 is, by Bayes' rule:

$$\begin{aligned} & \frac{\theta(u_1|\delta=1)\,\gamma}{\theta(u_1|\delta=1)\,\gamma + \theta(u_1|\delta=0)\,(1-\gamma)} \\ = & \frac{\varphi(u_1-1)\,\gamma}{\varphi(u_1-1)\,\gamma + \varphi(u_1)\,(1-\gamma)} \\ = & 1 \Big/ \left[1 + \frac{\varphi(u_1)(1-\gamma)}{\varphi(u_1-1)\gamma}\right]. \end{aligned} \tag{8}$$

By the monotone likelihood ratio property (MLRP) of θ, $\varphi(u_1)/\varphi(u_1 - 1)$ is decreasing in u_1 and so the whole expression is increasing in u_1. Let $\bar{u}$ be the utility level which solves (8) for 1/2: when $u_1 = \bar{u}$, the electorate is indifferent between a and b. For $u_1 > \bar{u}$, the electorate strictly prefers a and for $u_1 < \bar{u}$ it strictly prefers b. Algebra gives

$$\varphi(\bar{u})/\varphi(\bar{u} - 1) = \gamma/(1-\gamma); \tag{9}$$

this is at least 1, and as $\varphi(1/2)/\varphi(-1/2) = 1$ by symmetry, this shows that $\bar{u} \leqslant 1/2$. (9) has a unique solution as the left hand side is decreasing by the MLRP. In other words, the prior biases the electorate in favour of a; if $\gamma = 1/2$ the electorate is indifferent when $u_1 = 1/2$, which is equally likely whether d_1 is the right policy or not.

Write U for total expected utility over both periods, without discounting. Thus

$$U = \gamma + \gamma \Pr(d_2 = x | d_1 = x) + (1-\gamma) \Pr(d_2 = x | d_1 \neq x) \tag{10}$$

where the first term is expected utility in period 1, and the next two terms are expected utility in period 2 after respectively a correct and a wrong decision in period 1. We rewrite this as

$$\begin{aligned} U &= \gamma + \gamma \Pr(u_1 > \bar{u} | \delta = 1) + (1-\gamma) \Pr(u_1 < \bar{u} | \delta = 0) \\ &= \gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1-\gamma)\Phi(\bar{u}) \end{aligned} \tag{11}$$

□

Proposition 2. *When there is a single representative democracy:*

1. *There is a disciplined equilibrium if and only if* $\kappa \leqslant \Phi(\bar{u}) - \Phi(-\bar{u})$. *Voter utility in this equilibrium is* $1 + \pi$.

2. *There is a moral hazard equilibrium if and only if* $\kappa \geqslant \Phi(1 - \bar{u}) - \Phi(\bar{u} - 1)$. *Voter utility is* $2\pi + \pi(1-\pi)\{\Phi(\bar{u}) - \Phi(\bar{u} - 1)\}$.

3. *If and only if* $\Phi(\bar{u}) - \Phi(-\bar{u}) < \kappa < \Phi(1-\bar{u}) - \Phi(\bar{u}-1)$, *there is an equilibrium in mixed strategies: congruent types play* $d_1 = x$, *while non-congruent types play* $d_1 = a$ *when* $x = b$ *and mix between* a *and* b *when* $x = a$.

Proof. In all equilibria, because representatives always follow their preferences in period 2, the electorate rejects a representative iff its posterior belief that the representative is congruent is less than π.

In the disciplined equilibrium, we assume that a proportion ξ of representatives are impatient, and let $\xi \to 0$. These impatient types will always play $d_1 = r$, as for them $G/\beta(R+G) > 1$, that is, the benefit G of playing $d_1 = r$ outweighs the potential benefit $\beta(R+G)$ of staying in office, no matter what the probability of reelection is after playing $d_1 = r$ or $d_1 = x$. As, in the disciplined equilibrium, both congruent and non-congruent patient types play the same way, while only impatient non-congruent types deviate,

$$\begin{aligned}
\Pr(\text{congruent}|d_1 = a, u_1) &= \Pr(d_1 = a, u_1|\text{congruent})\Pr(\text{congruent})/ \\
&\quad [\Pr(d_1 = a,\, u_1|\text{congruent})\Pr(\text{congruent}) + \Pr(d_1 = a, \\
&\quad u_1|\text{non}-\text{congruent})\Pr(\text{non}-\text{congruent})] \\
&= \frac{\gamma\varphi(u_1-1)\pi}{(1-\xi)\gamma\varphi(u_1-1) + \xi(1-\gamma)\varphi(u_1)} \\
&= \frac{\pi}{(1-\xi) + \xi(1-\gamma)\varphi(u_1)/\gamma\varphi(u_1-1)} \qquad (12) \\
&\geqslant \pi \text{ iff } \frac{\varphi(u_1)}{\varphi(u_1-1)} \leqslant \frac{\gamma}{1-\gamma} \qquad (13)
\end{aligned}$$

and similarly

$$\Pr(\text{congruent}|d_1 = b,\, u_1) = \frac{\pi}{(1-\xi) + \xi\gamma\varphi(u_1)/(1-\gamma)\varphi(u_1-1)} \geqslant \pi \text{ iff } \frac{\varphi(u_1)}{\varphi(u_1-1)} \leqslant \frac{1-\gamma}{\gamma}. \qquad (14)$$

Let u_b solve $\frac{\varphi(u_b)}{\varphi(u_b-1)} = \frac{1-\gamma}{\gamma}$; similarly u_a solves $\frac{\varphi(u_a)}{\varphi(u_a-1)} = \frac{\gamma}{1-\gamma}$, hence $u_a = \bar{u} < 1/2 < u_b$ and indeed by symmetry of φ, $u_b = -(u_a - 1) = 1 - \bar{u}$. These are the utility cutpoints for reelection in a disciplined equilibrium after $d_1 = b$ and a respectively.

For a non-congruent representative to prefer not to deviate when $x = a$ we require

$$\begin{aligned} G+\beta(R+G)[1-\Phi(u_b)] &\leqslant \beta(R+G)[1-\Phi(u_a-1)] \\ \Leftrightarrow \kappa &\leqslant \Phi(u_b)-\Phi(u_a-1)=\Phi(1-\bar{u})-\Phi(\bar{u}-1), \end{aligned} \tag{15}$$

and similarly when $x = b$, we require

$$\kappa \leqslant \Phi(u_a)-\Phi(u_b-1)=\Phi(\bar{u})-\Phi(-\bar{u}) \tag{16}$$

which is a tighter condition as $\bar{u} < 1-\bar{u}$.

Examining (12) and (14) shows that the probability of an incumbent's congruence goes to π as $\xi \to 0$; since a new representative is also congruent with probability π, whether a representative is reelected or ejected makes no difference to utility in the limit. As representatives are disciplined in the first period but follow their preferences in period 2, utility is just $1 + \pi$. This completes the part of the proof relating to the disciplined equilibrium.

Turning to the moral hazard equilibrium, by Bayes' rule,

$$\Pr(\text{congruent}|d_1,u_1) = \frac{\Pr(d_1,u_1|\text{congruent})\,\pi}{\Pr(d_1,u_1|\text{congruent})\,\pi+\Pr(d_1,u_1|\text{non-cong.})\,(1-\pi)} \tag{17}$$

and

$$\begin{aligned} \Pr(d_1,u_1|\text{congruent}) &= \Pr(d_1|\text{congruent})\Pr(u_1|\text{congruent},d_1) \\ &= \Pr(x=d_1)\Pr(u_1|x=d_1) \\ &= \begin{cases} \gamma\,\varphi(u_1-1) & \text{if } d_1=a, \\ (1-\gamma)\,\varphi(u_1-1) & \text{if } d_1=b. \end{cases} \end{aligned} \tag{18}$$

Similarly

$$\Pr(d_1,u_1|\text{non}-\text{congruent}) = \begin{cases} (1-\gamma)\,\varphi(u_1-1) & \text{if } d_1=a, \\ \gamma\varphi(u_1-1) & \text{if } d_1=b \end{cases} \tag{19}$$

and thus

$$\Pr(\text{congruent}|d_1,u_1)=\begin{cases} \frac{\gamma\varphi(u_1-1)\pi}{\gamma\varphi(u_1-1)\pi+(1-\gamma)\varphi(u_1)(1-\pi)} & \text{for } d_1=a \\ \frac{(1-\gamma)\varphi(u_1-1)\pi}{(1-\gamma)\varphi(u_1-1)\pi+\gamma\varphi(u_1)(1-\pi)} & \text{for } d_1=b \end{cases} \tag{20}$$

where the second alternative is lower. The electorate reelects if and only if $\Pr(\text{congruent}|d_1,u_1) \geqslant \pi$. Exactly as before, $\bar{u}$ solves $\Pr(\text{congruent}|d_1,u_1)=\pi$ for u_1 when $d_1=a$, and $1-\bar{u}$ solves $\Pr(\text{congruent}|d_1,u_1)=\pi$ when $d_1=b$.

For a non-congruent type to shirk, she must prefer to harm her election prospects by choosing her own preferred action. The conditions are just the reverse of (14) and (16), i.e. $\kappa \geqslant \Phi(1-\bar{u}) - \Phi(\bar{u}-1)$ when $x = a$ and $\kappa \geqslant \Phi(\bar{u}) - \Phi(-\bar{u})$ when $x = b$; the former condition is tighter.

The electorate's expected utility in this equilibrium is, as in (4):

$$2\pi + \pi(1-\pi)(X - Y) \tag{21}$$

where

$$\begin{aligned} 1 - Y &\equiv \Pr(\text{reelected}|\text{congruent}) = \gamma[1 - \Phi(\bar{u}-1)] + (1-\gamma)[1 - \Phi(-\bar{u})] \\ \Leftrightarrow Y &= \gamma\Phi(\bar{u}-1) + (1-\gamma)\Phi(-\bar{u}) \end{aligned} \tag{22}$$

and

$$X \equiv \Pr(\text{ejected}|\text{non-congruent}) = \gamma\Phi(1-\bar{u}) + (1-\gamma)\Phi(\bar{u}) \tag{23}$$

Expected utility is thus

$$2\pi + \pi(1-\pi)\{\gamma[\Phi(1-\bar{u}) - \Phi(\bar{u}-1)] + (1-\gamma)[\Phi(\bar{u}) - \Phi(-\bar{u})]\} \tag{24}$$

Finally, we examine the mixed equilibrium. Given the specified strategies, by Bayes' rule

$$\Pr(\text{congruent}|d_1 = b, u_1) = \frac{\pi(1-\gamma)\varphi(u_1 - 1)}{\pi(1-\gamma)\varphi(u_1-1) + (1-\pi)\gamma(1-\alpha)\varphi(u_1)} \tag{25}$$

$$\text{and } \Pr(\text{congruent}|d_1 = a, u_1) = \frac{\pi\gamma\varphi(u_1-1)}{\pi\gamma\varphi(u_1-1) + (1-\pi)[\gamma\alpha\varphi(u_1-1) + (1-\gamma)\varphi(u_1)]}. \tag{26}$$

The voters reelect the incumbent iff these probabilities are at least π: write $\hat{u}_a$, $\hat{u}_b$ for the utility cutpoints, which solve

$$\frac{\varphi(\hat{u}_a)}{\varphi(\hat{u}_a - 1)} = \frac{\gamma(1-\alpha)}{1-\gamma} = \frac{\varphi(\hat{u}_b - 1)}{\varphi(\hat{u}_b)}. \tag{27}$$

By symmetry of φ, $\hat{u}_b = -(\hat{u}_a - 1)$. For the non-congruent a-type to mix actions, we require

$$\begin{aligned} G + \beta(R+G)[1 - \Phi(\hat{u}_b)] &= \beta(R+G)[1 - \Phi(\hat{u}_a - 1)] \\ \Leftrightarrow \kappa &= \Phi(\hat{u}_b) - \Phi(-\hat{u}_b). \end{aligned} \tag{28}$$

This equation determines $\hat{u}_b$, which in turn gives α in (27). Now $\frac{\varphi(\hat{u}_a)}{\varphi(\hat{u}_a - 1)} = \frac{\gamma(1-\alpha)}{1-\gamma} < \frac{\gamma}{1-\gamma} = \frac{\varphi(\bar{u})}{\varphi(\bar{u}-1)}$ implies $\hat{u}_a > \bar{u}$ and $\hat{u}_b < 1 - \bar{u}$. (28) then shows that $\kappa < \Phi(1-\bar{u}) - \Phi(\bar{u}-1)$.

The condition for the non-congruent b-type to play $d_1 = a$ is

$$\begin{aligned} G + \beta(R+G)[1-\Phi(\hat{u}_a)] &\geqslant \beta(R+G)[1-\Phi(\hat{u}_b - 1)] \\ \Leftrightarrow \kappa &\geqslant \Phi(\hat{u}_a) - \Phi(-\hat{u}_a) \end{aligned} \tag{29}$$

which implies $\kappa > \Phi(\bar{u}) - \Phi(-\bar{u})$.

As $\kappa \to \Phi(1-\bar{u}) - \Phi(\bar{u}-1)$, $\hat{u}_b \to 1 - \bar{u}$, hence $\hat{u}_a \to \bar{u}$ and by (27) and (9) we can see that $\alpha \to 0$. As $\kappa \to \Phi(\bar{u}) - \Phi(-\bar{u})$, $\hat{u}_b \to \bar{u}$ and by (27) and (9) we then have $\frac{\gamma(1-\alpha)}{1-\gamma} \to \frac{1-\gamma}{\gamma} \Leftrightarrow 1 - \alpha \to \frac{(1-\gamma)^2}{\gamma^2}$; this provides a lower bound on the probability of moral hazard in the mixed equilibrium.

Voter welfare is given by

$$\begin{aligned} &\gamma\pi\{1 + [1 - \Phi(\hat{u}_a - 1)] + \Phi(\hat{u}_a - 1)\pi\} \\ &+ \gamma(1-\pi)\{\alpha(1 + \Phi(\hat{u}_a - 1)\pi) + (1-\alpha)\Phi(\hat{u}_b)\pi\} \\ &+ (1-\gamma)\pi\{1 + [1 - \Phi(\hat{u}_b - 1)] + \Phi(\hat{u}_b - 1)\pi\} \\ &+ (1-\gamma)(1-\pi)\{\Phi(\hat{u}_a)\pi\} \end{aligned} \tag{30}$$

which simplifies to

$$2\pi + \pi(1-\pi)\{\gamma(1-\alpha)[\Phi(\hat{u}_b) - \Phi(-\hat{u}_b)] + (1-\gamma)[\Phi(\hat{u}_a) - \Phi(-\hat{u}_a)]\} + \gamma(1-\pi)\alpha. \tag{31}$$

As $\alpha \to 0$, $\hat{u}_a \to \bar{u}$ and this reduces to the utility of the moral hazard equilibrium (24). □

Proposition 3. *When there are $n > 1$ direct democratic states,*

1. *if and only if $\gamma < \bar{\gamma}$ where $\bar{\gamma}$ uniquely solves $\bar{\gamma} = \frac{1 + \Phi(-\bar{u})}{2 - \Phi(\bar{u}-1) + \Phi(-\bar{u})}$ there is a symmetric mixed equilibrium in which all electorates choose $d_1 = a$ with probability $\alpha \in [1/2, 1)$; for any fixed $\gamma < \bar{\gamma}$, α increases towards 1 as $n \to \infty$, and $\alpha^n \to \frac{2\gamma - 1}{\gamma\Phi(\bar{u}-1) + (1-\gamma)\Phi(-\bar{u})} > 0$ as $n \to \infty$.*

2. *otherwise, all states play* $d_1 = a$ *with certainty, giving expected voter utility of* $\gamma + \gamma(1 - \Phi(\bar{u} - 1)) + (1 - \gamma)\Phi(\bar{u}) < \gamma + 1$.

Proof. We seek a symmetric mixed equilibrium in which each electorate plays $d_1 = a$ with probability α. Given that other electorates are playing this strategy, each state's electorate gets utility from playing $d_1 = a$ of

$$\gamma + \gamma[\alpha^{n-1}(1 - \Phi(\bar{u} - 1)) + (1 - \alpha^{n-1})] + (1 - \gamma)[\alpha^{n-1}\Phi(\bar{u}) + (1 - \alpha^{n-1})] \tag{32}$$

where the terms in square brackets represent the probability of staying with a when $x = a$ and changing when $x = b$ respectively, and $\bar{u}$ is defined as in Proposition 1 to be the belief cutpoint after choosing $d_1 = a$. Similarly, each state gets utility from playing $d_1 = b$ of

$$(1 - \gamma) + (1 - \gamma)[1 - (1 - \alpha)^{n-1} + (1 - \alpha)^{n-1}(1 - \Phi(u_b - 1))] + \gamma[1 - (1 - \alpha)^{n-1} + (1 - \alpha)^{n-1}\Phi(u_b)] \tag{33}$$

where $u_b = 1 - \bar{u}$ is the belief cutpoint after choosing $d_1 = b$ which solves $\varphi(u_b)/\varphi(u_b -) = (1 - \gamma)/\gamma$, and $(1 - \alpha)^{n-1}$ is the probability of all other states choosing b. Mixing requires that (32) and (33) are equal: we simplify both sides, using $1 - \Phi(\bar{u}) = \Phi(-\bar{u})$ and $1 - \Phi(u_b) = \Phi(\bar{u} - 1)$, to get

$$\begin{aligned}
\gamma + 1 - \alpha^{n-1}[\gamma\Phi(\bar{u} - 1) + (1 - \gamma)\Phi(-\bar{u})] &= (1 - \gamma) + 1 - (1 - \alpha)^{n-1}[\gamma\Phi(\bar{u} - 1) + (1 - \gamma)\Phi(-\bar{u})] \\
\Leftrightarrow 2\gamma - 1 &= [\alpha^{n-1} - (1 - \alpha)^{n-1}][\gamma\Phi(\bar{u} - 1) + (1 - \gamma)\Phi(-\bar{u})] \\
\Leftrightarrow \alpha^{n-1} - (1 - \alpha)^{n-1} &= \frac{2\gamma - 1}{\gamma\Phi(\bar{u} - 1) + (1 - \gamma)\Phi(-\bar{u})}. \qquad (34)
\end{aligned}$$

As the left hand side is less than 1 for $\alpha < 1$, there is a non-degenerate mixed equilibrium if and only if

$$\begin{aligned}
2\gamma - 1 &< \gamma\Phi(\bar{u} - 1) + (1 - \gamma)\Phi(-\bar{u}) \qquad (35) \\
\Leftrightarrow \gamma &< \frac{1 + \Phi(-\bar{u})}{2 - \Phi(\bar{u} - 1) + \Phi(-\bar{u})}; \qquad (36)
\end{aligned}$$

otherwise (33) is always strictly less than (32), as can be seen by taking $\alpha \to 1$. The derivative of the right hand side of (35) with respect to γ, recalling that $\bar{u}$ depends on γ, is

$$\begin{aligned}
& \Phi(\bar{u}-1)-\Phi(-\bar{u})+\gamma\frac{d}{d\gamma}\Phi(\bar{u}-1)+(1-\gamma)\frac{d}{d\gamma}\Phi(-\bar{u}) \\
= & \Phi(\bar{u}-1)-\Phi(-\bar{u})+\gamma\varphi(\bar{u}-1)\frac{d\bar{u}}{d\gamma}+(1-\gamma)\varphi(-\bar{u})\left(-\frac{d\bar{u}}{d\gamma}\right) \\
= & \Phi(\bar{u}-1)-\Phi(-\bar{u})+(1-\gamma)\varphi(\bar{u})\frac{d\bar{u}}{d\gamma}-(1-\gamma)\varphi(\bar{u})\frac{d\bar{u}}{d\gamma}, \\
& \text{where we use that } \varphi(\bar{u})=\varphi(-\bar{u}) \text{ and } \varphi(\bar{u})/\varphi(\bar{u}-1)=\gamma/(1-\gamma), \\
= & \Phi(\bar{u}-1)-\Phi(-\bar{u}) \qquad (37) \\
< & 0 \text{ as } \bar{u}<1/2.
\end{aligned}$$

On the other hand, the derivative of the left hand side of (35) with respect to γ is positive. Thus there is a unique cutpoint $\bar{\gamma}$ solving (35) with equality, and we have a mixed strategy equilibrium for $\gamma < \bar{\gamma}$ and a pure strategy equilibrium for $\gamma \geqslant \bar{\gamma}$. When $\gamma > \bar{\gamma}$, all states play $d_1 = a$, and therefore voters gain no more information than in the case of a single direct democracy: thus all voters update identically after period 1, using the same cutpoint $\bar{u}$ as in Proposition 1, and expected utility is $\gamma+\gamma(1-\Phi(\bar{u}-1))+(1-\gamma)\Phi(\bar{u})$ just as before.

Finally, examining (34) as $n \to \infty$, because the right hand side is constant with respect to n, $\alpha^{n-1}-(1-\alpha)^{n-1}$ remains so too. This and $\alpha > 1/2$ require that α increase with n. Indeed, $\alpha \to 1$ as $n \to \infty$. For suppose α is bounded above by $K<1$. Then for $m>n$, $\alpha^{m-1}-(1-\alpha)^{m-1} < K^{m-n}\alpha^{n-1}-(1-K)^{m-n}(1-\alpha)^{n-1}$ and as $m \to \infty$ this goes to 0, contradicting the fact that the right hand side of (34) is positive. As α is increasing and not bounded by $K<1$ it must be that it approaches 1. Therefore as $n \to \infty$, $1-\alpha \to 0$ and $(1-\alpha)^{n-1} < 1-\alpha$ also goes to 0. Thus, $\alpha^{n-1}-(1-\alpha)^{n-1} \to \alpha^{n-1}$. Thus α^{n-1} approaches the right hand side of (34), and α^n does the same. □

Proposition 6. *1. For $\kappa < \Phi(\bar{u})$ and any $\gamma > \bar{\gamma}$, there is some π such that neither a system of representative democracies nor a system of direct democracies is stable.*

2. For $\kappa > 1$ and any $\gamma > \bar{\gamma}$, there is some π such that when n is large enough, neither a system of n representative democracies nor a system of n direct democracies is stable.

Proof. By Proposition 4, if $\kappa \leqslant 1$ then 2 or more representative democracies can achieve voter utility of $1+\pi$ in a disciplined equilibrium, and if $\kappa > 1$ and $n \to \infty$ then representative democracies achieve voter utility approaching $2\pi + \pi(1-\pi)$ in a moral hazard equilibrium. In either case, as $n \to \infty$, a single direct democracy among $n-1$ representative democracies will achieve utility approaching $1+\gamma$ (Proposition 5). So for a system of representative democracies to be stable it is necessary that

$$1+\pi \geqslant 1+\gamma \quad \text{for } \kappa \leqslant 1; \text{ or} \tag{38}$$

$$2\pi + \pi(1-\pi) \geqslant 1+\gamma \quad \text{for } \kappa > 1. \tag{39}$$

If one of these conditions is violated, then (for high enough n, when $\kappa \geqslant 1$) a system of n representative democracies will be "invaded", in the evolutionary sense, by a single direct democracy.

From Proposition 3, if $\gamma \geqslant \bar{\gamma}$, then in a system of n direct democracies, all electorates choose $d_1 = a$ and achieve utility of $\gamma + \gamma(1-\Phi(\bar{u}-1)) + (1-\gamma)\Phi(\bar{u})$. The condition for stability of a direct democratic federation is thus

$$\gamma[2-\Phi(\bar{u}-1)] + (1-\gamma)\Phi(\bar{u}) \geqslant U' \tag{40}$$

where U' stands for the utility achieved by a single representative democracy, say state 1, with $n-1$ direct democratic neighbours.

First, suppose the representative in state 1 is disciplined. Then the direct democracies will all choose $d_1^j = a$, $j \in \{2, \ldots, n\}$, and follow the representative democracy in period 2. To analyse the conditions for this equilibrium we assume that a small proportion ξ of representatives are impatient. If the electorate observes $d_1^1 = b$ and $u_1^1 < u_1^2$, given equilibrium strategies it will infer that its representative is impatient and non-congruent, and eject him. Observing $d_1^1 = b$ and

$u_1^1 > u_1^2$ will give a posterior probability just above π (due to the small proportion of impatient, congruent representatives) and the electorate will reelect.

If $d_1^1 = a$ then by Bayes' rule

$$\Pr(\text{congruent}|d_1^1 = a, u_1^1) = \frac{\pi\varphi(u_1^1 - 1)\gamma}{\pi\varphi(u_1^1 - 1)\gamma + (1-\pi)\{\varphi(u_1^1 - 1)\gamma(1-\xi) + \varphi(u_1^1)(1-\gamma)\xi\}} \tag{41}$$

where the two terms in curly brackets reflect the probabilities of patient and impatient non-congruent representatives. The condition for this to equal π is

$$\begin{aligned}
\pi &= \frac{\pi\varphi(u_1^1 - 1)\gamma}{\pi\varphi(u_1^1 - 1)\gamma + (1-\pi)\{\varphi(u_1^1 - 1)\gamma(1-\xi) + \varphi(u_1^1)(1-\gamma)\xi\}} \\
\Leftrightarrow \varphi(u_1^1 - 1)\gamma &= \pi\varphi(u_1^1 - 1)\gamma + (1-\pi)\{\varphi(u_1^1 - 1)\gamma(1-\xi) + \varphi(u_1^1)(1-\gamma)\xi\} \\
\Leftrightarrow \varphi(u_1^1 - 1)\gamma &= \varphi(u_1^1 - 1)\gamma(1-\xi) + \varphi(u_1^1)(1-\gamma)\xi \\
\Leftrightarrow \frac{\varphi(u_1^1)}{\varphi(u_1^1 - 1)} &= \frac{\gamma}{1-\gamma} \\
\Leftrightarrow u_1^1 &= \bar{u}
\end{aligned} \tag{42}$$

, so $\bar{u}$ is the state 1 electorate's cutpoint for reelection when $d_1^1 = a$. Conditions for discipline when $x = a$ and when $x = b$ are therefore

$$G \leqslant \beta(R+G)[1 - \Phi(\bar{u} - 1)] \Leftrightarrow \kappa \leqslant \Phi(1 - \bar{u}) \tag{43}$$

and

$$G + \beta(R+G)[1 - \Phi(\bar{u})] \leqslant \beta(R+G) \Leftrightarrow \kappa \leqslant \Phi(\bar{u}) \tag{44}$$

respectively, reflecting the certainty of reelection (ejection) from choosing $d_1^1 = b = x$ ($d_1^1 = b \neq x$). The latter condition is tighter, as $\bar{u} < 1 - \bar{u}$. Thus, if $\kappa \leqslant \Phi(\bar{u})$, there is a disciplined equilibrium and $U' = 1 + \pi$. Writing $\gamma[2 - \Phi(\bar{u} - 1)] + (1 - \gamma)\Phi(\bar{u}) = \gamma + \gamma\Phi(1 - \bar{u}) + (1-\gamma)\Phi(\bar{u})$, and $1 + \gamma = \gamma + \gamma + (1-\gamma)$, we see that when $\gamma \geqslant \bar{\gamma}$ and

$$\gamma + \gamma\Phi(1 - \bar{u}) + (1-\gamma)\Phi(\bar{u}) < 1 + \pi < \gamma + \gamma + (1-\gamma), \tag{45}$$

(40) and (38) will simultaneously be violated. Clearly for any $\gamma \geqslant \bar{\gamma}$ there will be values of π satisfying this pair of inequalities. In these cases only a mixed federation of direct and representative democracies can be stable. This proves part 1.

If $\kappa > 1$ then a single representative democracy will always be in a moral hazard equilibrium. The extra information provided by the single representative state provides an added incentive for its direct democratic neighbours to follow their prior; they would do so anyway as $\gamma \geqslant \bar{\gamma}$. So $d_1^j = a$, for $j = 2, \ldots, n$. Therefore, a non-congruent a-type who plays $d_1^1 = b$ is detected with certainty, as is a congruent b-type who plays $d_1^1 = b$. A representative who plays $d_1^1 = a$ may be ejected if u_1^1 is too low; the presence of direct democratic neighbours gives no extra information in this case and so the utility cutpoint for ejection is just the same as in the unitary case: $\bar{u}$ solving (20) when $d_1 = a$. Expected utility when $x = a$ is thus

$$\pi[1 + (1 - \Phi(\bar{u} - 1)) + \pi\Phi(\bar{u} - 1)] + (1 - \pi)\pi \tag{46}$$

where the first term inside the square brackets is period 1 utility from a congruent representative, the second is period 2 utility if the congruent incumbent is reelected and the third term is period 2 utility if the congruent incumbent is replaced; and the final term outside the square brackets represents expected period 2 utility after a non-congruent incumbent is certainly replaced. Expected utility when $x = b$ is $2\pi + (1 - \pi)\pi\Phi(\bar{u})$ by similar logic, and so

$$\begin{aligned} U' &= \gamma\{\pi[1 + (1 - \Phi(\bar{u} - 1)) + \pi\Phi(\bar{u} - 1)] + (1 - \pi)\pi\} + (1 - \gamma)\{2\pi + (1 - \pi)\pi\Phi(\bar{u})\} \\ &= \pi\{\gamma[1 + (1 - \Phi(\bar{u} - 1)) + \pi\Phi(\bar{u} - 1) + (1 - \pi)] + (1 - \gamma)[2 + (1 - \pi)\Phi(\bar{u})]\} \\ &= \pi\{2 - \gamma(1 - \pi)\Phi(\bar{u} - 1) + (1 - \gamma)(1 - \pi)\Phi(\bar{u})\} \\ &= 2\pi + \pi(1 - \pi)[\gamma\Phi(1 - \bar{u}) + (1 - \gamma)\Phi(\bar{u})]. \end{aligned} \tag{47}$$

Thus (40) will be violated iff

$$\begin{aligned} 2\pi + \pi(1 - \pi)[\gamma\Phi(1 - \bar{u}) + (1 - \gamma)\Phi(\bar{u})] &> \gamma[2 - \Phi(\bar{u} - 1)] + (1 - \gamma)\Phi(\bar{u}) \\ &= \gamma + \gamma\Phi(1 - \bar{u}) + (1 - \gamma)\Phi(\bar{u}) \\ \Leftrightarrow 2\pi + [\pi(1 - \pi) - 1][\gamma\Phi(1 - \bar{u}) + (1 - \gamma)\Phi(\bar{u})] &> \gamma \\ \Leftrightarrow 2\pi + \pi(1 - \pi)[\gamma\Phi(1 - \bar{u}) + (1 - \gamma)\Phi(\bar{u})] & \\ + 1 - [\gamma\Phi(1 - \bar{u}) + (1 - \gamma)\Phi(\bar{u})] &> 1 + \gamma \end{aligned} \tag{48}$$

and so (39) and (40) will simultaneously be violated iff

$$2\pi + \pi(1-\pi)[\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] + 1 - [\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] > 1 + \gamma > 2\pi + \pi(1-\pi). \tag{49}$$

For any γ, we can choose π so that $2\pi + \pi(1-\pi)$ approaches $1+\gamma$ from below, thus ensuring the second inequality in the chain; it remains to prove that

$$\begin{aligned}
2\pi + \pi(1-\pi)[\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] & \\
+1 - [\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] &> 2\pi + \pi(1-\pi) \\
\Leftrightarrow \pi(1-\pi)[\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] & \\
+1 - [\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] &> \pi(1-\pi) \\
\Leftrightarrow 1 - [\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] &> \pi(1-\pi)\{1 - \\
& \quad [\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)]\} \\
\Leftrightarrow 1 &> \pi(1-\pi)
\end{aligned} \tag{50}$$

which is always the case.

Thus for $\kappa > 1$ and any γ, when $2\pi + \pi(1-\pi)$ is close enough to $1+\gamma$, but below it, (39) and (40) are violated and for high enough n neither a system of n representative democracies nor one of n direct democracies is stable. This proves part 2. □

Remark. For values of κ between $\Phi(\bar{u}\,)$ and 1, and some values of γ, it may be that one of the two systems is stable whatever the value of π. For example, suppose $\kappa = 1$. Then as a system of representative democracies will have a disciplined equilibrium, it will be stable if (38) is satisfied. A single RD will have a moral hazard equilibrium: a non-congruent a-type will be ejected if she chooses $d_1^1 = b$, but may not be reelected if she chooses $d_1^1 = a$, and so will prefer the former; a non-congruent b-type will be reelected for sure if she chooses $d_1^1 = b$ but may get away with choosing $d_1^1 = a$, and so will prefer the latter. The voters' expected utility will be (47) as before. Therefore (38) and (40) are violated, by analogy with (49), if and only if

$$2\pi + \pi(1-\pi)[\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] + 1 - [\gamma\Phi(1-\bar{u}\,) + (1-\gamma)\Phi(\bar{u}\,)] > 1 + \gamma > 1 + \pi \tag{51}$$

The left hand side is increasing in π and so both inequalities will hold simultaneously if and only if they hold when $\pi \nearrow \gamma$, thus iff

$$\begin{aligned} 2\gamma + \gamma(1-\gamma)[\gamma\Phi(1-\bar{u}) + (1-\gamma)\Phi(\bar{u})] + 1 - [\gamma\Phi(1-\bar{u}) + (1-\gamma)\Phi(\bar{u})] &> 1+\gamma \\ \Leftrightarrow \gamma + [\gamma(1-\gamma)-1][\gamma\Phi(1-\bar{u}) + (1-\gamma)\Phi(\bar{u})] &> 0 \end{aligned} \quad (52)$$

Suppose that the voters' utility signal becomes very accurate, so $\Phi(\varepsilon) \to 1$ for $\varepsilon > 0$. Then if $\bar{u} \in (0, 1/2)$, the left hand side above approaches $\gamma + [\gamma(1-\gamma) - 1] = 2\gamma - \gamma^2 - 1 = -(1-\gamma)^2$ and thus cannot be positive. And when the voters' utility signal is very accurate, $\bar{u}$ will indeed be close to 1/2 as e.g. $\bar{u}$ a little larger than 1/2 will provide very good evidence that the right policy was chosen. For example, if ε is normally distributed with variance σ^2 then simple algebra from the pdf shows that $\bar{u} = 1/2 - \sigma^2 \log(\gamma/(1-\gamma))$ which will be positive for low enough σ^2. Thus when the voters' signal is accurate enough, there is no way for (38) and (40) to be violated simultaneously.

Nevertheless, there will also be cases when (52) holds: for example if ε is normally distributed and σ^2 is large then (52) approaches $\gamma + [\gamma(1-\gamma) - 1][\gamma] = \gamma^2(1-\gamma) > 0$. So there are cases in which only a mixed system of RDs and DDs will be stable, but no simple general rule.

Proposition 9. *If $\kappa > 1$ and n is large, and $\gamma > \bar{\gamma}$ as defined in Proposition 3; then*

1. *when $2\pi + \pi(1-\pi) > 1 + \gamma$ then a system of all representative democracies is stable and maximizes voter welfare;*

2. *when $2\pi + \pi(1-\pi) < 1 + \gamma$ then in general the equilibrium number of representative democracies is smaller than the welfare-maximizing number of representative democracies.*

Proof. By (7) we can see that if $2\pi + \pi(1-\pi) > 1 + \gamma$ then for n large enough, it will be best to have an all-representative federation, and this federation will be stable as any RD would lose utility by switching to a DD. This proves part 1.

Suppose that there are $m < n$ RDs and $n - m$ DDs. As $\gamma > \bar{\gamma}$ all DDs will choose $d_1 = a$: the possibility that one of the RDs will choose b and do their experimenting for them only increases the incentive to choose a so the argument from Proposition 3 applies *a fortiori*. As $\kappa > 1$ all non-congruent representatives shirk. Thus when $x = a$, the probability that all states choose $d_1 = a$ is π^m, and when $x = b$, the probability that all states choose $d_1 = a$ is $(1-\pi)^m$; otherwise at least one state chooses $d_1 = b$.

Expected voter utility in an RD is

$$\begin{aligned} & \gamma\{\pi[2-(1-\pi)\Pr(\text{ejected}|d_1=a=x)] + (1-\pi)\pi\Pr(\text{ejected}|d_1=b\neq x)\} + (1-\gamma)\{\pi[2-(1-\pi)\Pr(\text{ejected}|d_1=b=x)] + (1-\pi)\pi\Pr(\text{ejected}|d_1=a\neq x)\} \\ = & \gamma\{\pi[2-(1-\pi)\Pr(\text{ejected}|d_1=a=x)] + (1-\pi)\pi\} + (1-\gamma)\{2\pi + (1-\pi)\pi\Pr(\text{ejected}|d_1=a\neq x)\} \qquad (53) \end{aligned}$$

as playing b results in the electorate learning the representative's congruence with certainty. Let $u_a(m)$ be the cutpoint for reelection when all states, including m RDs, choose $d_1 = a$.

$$\begin{aligned} \Pr(\text{ejected}|d_1=a=x) &= \pi^{m-1}\Phi(u_a(m)-1) \qquad (54) \\ \Pr(\text{ejected}|d_1=a\neq x) &= (1-\pi)^{m-1}\Phi(u_a(m)) + (1-(1-\pi)^{m-1}) \qquad (55) \\ &= 1-(1-\pi)^{m-1}\Phi(-u_a(m)) \end{aligned}$$

as when at least one state from the remaining $m-1$ RDs chooses $d_1 = b$, the electorate learns the representative's congruence with certainty. Plugging these into (53) gives expected voter utility in an RD of

$$\begin{aligned} U_{\text{RD}}(m) &= \gamma\{\pi[2-(1-\pi)\pi^{m-1}\Phi(u_a(m)-1)] + (1-\pi)\pi\} + (1-\gamma)\{2\pi + (1-\pi)\pi[1-(1-\pi)^{m-1}\Phi(-u_a(m))]\} \\ &= 2\pi + \pi(1-\pi) - \gamma(1-\pi)\pi^m\Phi(u_a(m)-1) - (1-\gamma)\pi(1-\pi)^m\Phi(-u_a(m)). \qquad (56) \end{aligned}$$

Expected voter utility in a DD is

$$\begin{aligned}U_{\mathrm{DD}}(m) &= \gamma+\gamma(1-\pi^m)+\gamma\pi^m[1-\Phi(\hat{u}_a(m)-1)]+(1-\gamma)(1-(1-\pi)^m)+(1-\\&\quad\gamma)(1-\pi)^m\Phi(\hat{u}_a(m))\\&= \gamma+1-\gamma\pi^m\Phi(\hat{u}_a(m)-1)-(1-\gamma)(1-\pi)^m\Phi(-\hat{u}_a(m))\end{aligned} \tag{57}$$

where $\hat{u}_a(m)$ gives the voters' cutpoint for choosing $d_2=a$ when all states, including m representative democracies, choose $d_1=a$.

Differentiating U_{DD} with respect to m, we note that

$$\frac{d}{dm}U_{\mathrm{DD}}=\frac{\partial}{\partial m}U_{\mathrm{DD}}+\frac{\partial}{\partial\hat{u}_a}U_{\mathrm{DD}}\cdot\frac{d\,\hat{u}_a}{dm} \tag{58}$$

and as $\hat{u}_a$ is the welfare-maximizing choice, we can apply the envelope theorem to conclude

$$\frac{d}{dm}U_{\mathrm{DD}}=\frac{\partial}{\partial m}U_{\mathrm{DD}}=-\gamma\pi^m(\ln\pi)\,\Phi(\hat{u}_a(m)-1)-(1-\gamma)(1-\pi)^m(\ln(1-\pi))\Phi(-\hat{u}_a(m))>0. \tag{59}$$

Similarly,

$$\frac{d}{dm}U_{\mathrm{RD}}=\frac{\partial}{\partial m}U_{\mathrm{RD}}=-\gamma(1-\pi)\pi^m\Phi(u_a(m)-1)\ln\pi-(1-\gamma)\pi(1-\pi)^m\Phi(-u_a(m))\ln(1-\pi)>0. \tag{60}$$

Expected average utility is

$$\frac{m}{n}U_{\mathrm{RD}}(m)+\frac{n-m}{n}U_{\mathrm{DD}}(m). \tag{61}$$

Let m^* be the number of RDs that maximizes this. I show that $m^*\geqslant\hat{m}$.

To be in equilibrium, $\hat{m}$ must satisfy

$$\begin{aligned}U_{\mathrm{DD}}(\hat{m})-U_{\mathrm{RD}}(\hat{m}+1) &\geqslant 0\\ U_{\mathrm{RD}}(\hat{m})-U_{\mathrm{DD}}(\hat{m}-1) &\geqslant 0\end{aligned} \tag{62}$$

so that no state wishes to change systems. Write $\Delta(m)=U_{\mathrm{DD}}(m)-U_{\mathrm{RD}}(m+1)$. I next show $\Delta(m)$ is increasing.

$$\begin{aligned}\frac{d}{dm}\Delta(m) &= U'_{\mathrm{DD}}(m)-U'_{\mathrm{RD}}(m+1)\\&= -\gamma\pi^m(\ln\pi)\,\Phi(\hat{u}_a(m)-1)-(1-\gamma)(1-\pi)^m(\ln(1-\pi))\Phi(-\hat{u}_a(m))\\&\quad-\{-\gamma(1-\pi)\pi^m\Phi(u_a(m+1)-1)\ln\pi-(1-\gamma)\pi(1-\pi)^m\Phi(-\\&\quad u_a(m+1))\ln(1-\pi)\}\\&= -\gamma\pi^m(\ln\pi)[\Phi(\hat{u}_a(m)-1)-(1-\pi)\Phi(u_a(m+1)-1)]\\&\quad-(1-\gamma)(1-\pi)^m(\ln(1-\pi))[\Phi(-\hat{u}_a(m))-\pi\Phi(-u_a(m+1))]\end{aligned} \tag{63}$$

Now, the RD reelection cutpoint $u_a(m)$ solves

$$\begin{aligned} \pi &= \Pr(x=a|u_a, d_1^i = a \forall i) \\ &= \frac{\varphi(u_a-1)\pi^m\gamma}{\varphi(u_a-1)\pi^m\gamma+\varphi(u_a)(1-\pi)^m(1-\gamma)} \\ \Leftrightarrow \frac{\varphi(u_a)}{\varphi(u_a-1)} &= \frac{\gamma}{1-\gamma}\frac{\pi^{m-1}}{(1-\pi)^{m-1}} \end{aligned} \tag{64}$$

while the DD cutpoint for choosing a, $\hat{u}_a(m)$, solves

$$\begin{aligned} \frac{1}{2} &= \Pr(x=a|\hat{u}_a, d_1^i = a \forall i) \\ &= \frac{\varphi(\hat{u}_a-1)\pi^m\gamma}{\varphi(\hat{u}_a-1)\pi^m\gamma+\varphi(\hat{u}_a)(1-\pi)^m(1-\gamma)} \\ \Leftrightarrow \frac{\varphi(\hat{u}_a)}{\varphi(\hat{u}_a-1)} &= \frac{\gamma}{1-\gamma}\frac{\pi^{m}}{(1-\pi)^{m}}. \end{aligned} \tag{65}$$

Comparing these results shows that $\hat{u}_a(m) = u_a(m+1)$. Plugging this identity into (63) shows that $\Delta(m)$ is increasing.

(62) shows $\Delta(\hat{m}-1) \leqslant 0$. As $\Delta(m)$ is increasing, $\Delta(m) \leqslant 0$ for $m \leqslant \hat{m}-1$. Suppose $m^* \leqslant \hat{m}-1$. Now average utility at m^* is at least as high as at m^*+1:

$$\frac{m^*}{n}U_{\mathrm{RD}}(m^*) + \frac{n-m^*}{n}U_{\mathrm{DD}}(m^*) \geqslant \frac{m^*+1}{n}U_{\mathrm{RD}}(m^*+1) + \frac{n-m^*-1}{n}U_{\mathrm{DD}}(m^*+1)$$

or, rearranging:

$$\frac{m^*}{n}[U_{\mathrm{RD}}(m^*) - U_{\mathrm{RD}}(m^*+1)] + \frac{n-m^*}{n}[U_{\mathrm{DD}}(m^*) - U_{\mathrm{DD}}(m^*+1)] + \frac{1}{n}[U_{\mathrm{DD}}(m^*+1) - U_{\mathrm{RD}}(m^*+1)] \geqslant 0$$

and since the first two terms are negative, by state utility increasing in m, it must be that

$$U_{\mathrm{DD}}(m^*+1) - U_{\mathrm{RD}}(m^*+1) > 0, \tag{66}$$

hence, again by utility increasing in m,

$$U_{\mathrm{DD}}(m^*+1) - U_{\mathrm{RD}}(m^*) > 0. \tag{67}$$

But

$$\Delta(m^*) = U_{\mathrm{DD}}(m^*) - U_{\mathrm{RD}}(m^*+1) \leqslant 0 \tag{68}$$

by $m^* \leqslant \hat{m}-1$. Deducting these gives

$$[U_{\mathrm{DD}}(m^*) - U_{\mathrm{DD}}(m^*+1)] + [U_{\mathrm{RD}}(m^*) - U_{\mathrm{RD}}(m^*+1)] < 0, \tag{69}$$

a contradiction by utility increasing in m. Thus $m^* \geqslant \hat{m}$.

It is possible that $m^* = \hat{m}$ for n small. In general, however, we would expect $\hat{m} > m^*$ for large values of n. This can be seen intuitively by looking at the continuous equivalents of our optimality and equilibrium conditions. For optimality, the FOC is

$$\frac{m}{n}U'_{\text{RD}}(m^*) + \frac{n-m^*}{n}U'_{\text{DD}}(m^*) + \frac{1}{n}[U_{\text{RD}}(m^*) - U_{\text{DD}}(m^*)] = 0 \tag{70}$$

and as the first two terms are positive, the last term must be negative; but if so a representative democracy at m^* will have an incentive to switch to a direct democracy. □

Notes

1. I use "states" throughout to refer indifferently to political units with policy-making power. These could be nation-states or regional sub-units such as Swiss cantons.
2. The conclusion suggests a mechanism by which direct democracies may have a larger impact on their neighbours.
3. Adding a discount rate for voters would alter decisions only in a direct democracy, as in a representative democracy the voters never make a decision before the end of period 1. The main change would be a greater incentive to choose $d_1 = a$ when there are many DD states (Section 4.3).
4. For $\gamma < \bar{\gamma}$, there are also n pure strategy equilibria in which one state chooses $d_1 = b$ and all others choose $d_1 = a$. The proof is simple and available on request.
5. Weakening this assumption would not affect the limiting result in Proposition 4 as $n \to \infty$, so long as representatives' congruence is not perfectly correlated among all the states.

6. The same is true if π is high. For example if $\pi = 0.95$, $1 - 0.95^{20} - (1 - 0.95)^{20} \cong 0.6$, so even with 20 states shirking will only be detected 60% of the time.

7. I do not look for a mixed equilibrium of constitutional choice, as I do in the case of policy choice, because it is hard to interpret a mixed strategy profile within the historical interpretation of equilibrium.

Bibliography

Aghion, P., A. F. Alesina, and F. Trebbi (2002). "Endogenous Political Institutions." NBER Working Paper 9006.

Arceneaux, K. (2002). "Direct Democracy and the Link between Public Opinion and State Abortion Policy." State Politics and Policy Quarterly 2 (4):372-87.

Bailey, M. A., M. C. Rom, and M. M. Taylor. 2004. "State competition in higher education: A race to the top, or a race to the bottom?" Economics of Governance 5 (1):53-75.

Barbera, S., and M. O. Jackson (2004). "Choosing How to Choose: Self-Stable Majority Rules and constitutions." Quarterly Journal of Economics 119 (3):1011-1048.

Barro, R. J. (1973). "The control of politicians: An economic model." Public Choice 14 (1):19-42.

Berry, F. S., and W. D. Berry (1999). "Innovation and Diffusion Models in Policy Research." Theories of the Policy Process:169-200.

Besley, T., and S. Coate (2003). "Centralized versus decentralized provision of local public goods: a political economy approach." Journal of Public Economics 87 (12):2611-37.

Boehmke, F. J. (1999). "Casino Gaming Diffusion, the Initiative Process and Expected Voter Support." In Southern Political Science Association.

Boix, C. (2003). Democracy and Redistribution. Cambridge: Cambridge University Press.

Bowler, S., and T. Donovan (2004). "Measuring the Effects of Direct Democracy on State Policy: Not All Initiatives Are Created Equal." State Politics and Policy Quarterly 4 (3):345-63.

Burden, B. C. (2005). "Institutional Features and Policy Representation in the States." State Politics and Policy Quarterly 5:373-93.

California Healthcare Foundation (2005). "Background: Similarities & Differences between Competing Propositions 78 & 79."
<http://www.healthvote.org/index.php/site/article/C33/79_background>.
Accessed 3 September 2007.

Camobreco, J. F. (1998). "Preferences, Fiscal Policies, and the Initiative Process." Journal of Politics 60(3):819-29.

Cronin, T. E. (1989). Direct Democracy: The Politics of Initiative, Referendum, and Recall. Cambridge, Mass.: Harvard University Press.

Feld, L. P., and M. Savioz (1997) "Direct Democracy Matters for Economic Performance: An Empirical Investigation." Kyklos 50:507-38.

Feld, L. P., and G. Kirchgässner (2000). "Direct Democracy, Political Culture and the Outcome of Economic Policy: A Report on the Swiss Experience." European Journal of Political Economy 16 (2):287-306.

Ferejohn, J. (1986). "Incumbent performance and electoral control." Public Choice 50 (1):5-25.

Frey, B. S., and A. Stutzer. (2000). "Happiness, Economy and Institutions." The Economic Journal 110 (466):918-38.

Gerber, E. R. (1996). "Legislative Response to the Threat of Popular Initiatives." American Journal of Political Science 40 (1):99-128.

Gerber, E. R. (1999). The Populist Paradox: Interest Group Influence and the Promise of Direct Legislation. New Jersey: Princeton University Press.

Haider-Markel, D. P. (2001). “Policy Diffusion as a Geographical Expansion of the Scope of Political Conflict: Same-Sex Marriage Bans in the 1990s.” State Politics and Policy Quarterly 1 (1):5-26.

Huntington, S. P. (1991). The third wave. Norman, Oklahoma: University of Oklahoma Press.

Kirlin, J. J. (1982). “The impacts of Proposition 13 upon the California political system: Re-regulating the intergovernmental system.” Public Choice 39 (1):147-69.

Lascher, E. L. L., Jr, M. G. Hagen, and S. A. Rochlin (1996). “Gun Behind the Door? Ballot Initiatives, State Policies and Public Opinion.” Journal of Politics 58 (3):760-75.

Maskin, Eric and Jean Tirole (2004). “The Politician and the Judge: Accountability in Government.” American Economic Review 94 (4):1034-1054.

Matsusaka, John G. (2001). “Problems with a Methodology Used to Evaluate the Voter Initiative.” Journal of Politics 63 (4):1250-6.

Matsusaka, John G. (2004). For the Many Or the Few: The Initiative, Public Policy, and American Democracy. Chicago: University of Chicago Press.

Oates, W. E. (1999). “An Essay on Fiscal Federalism.” Journal of Economic Literature 37 (3):1120-49.

Rincke, J. (2007). “Policy diffusion in space and time: The case of charter schools in California school districts.” Regional Science and Urban Economics 37 (5):526-41.

Rose-Ackerman, Susan (1980). “Risk-taking and Reelection: Does Federalism Promote Innovation?” Journal of Legal Studies 9:593-616.

Salmon, P. (1987). “Decentralization as an Incentive Scheme.” Oxford Review of Economic Policy 3 (2):24.

Schaltegger, C. A., and D. Küttel (2002). “Exit, Voice, and Mimicking Behavior: Evidence from Swiss Cantons.” Public Choice 113 (1):1-23.

Smith, D. A. (1998). Tax Crusaders and the Politics of Direct Democracy. London: Routledge.

Strumpf, Koleman S. (2002). "Does Government Decentralization Increase Policy Innovation?" Journal of Public Economic Theory 4:207-241.

Walker, J. L. (1969). "The Diffusion of Innovations among the American States." American Political Science Review 63 (3):880-99.

Wrede, M. (2001). "Yardstick competition to tame the Leviathan." European Journal of Political Economy 17 (4):705-21.