Online Appendix
Democracy and its Vulnerabilities: Dynamics of Democratic Backsliding

Formal Proofs

Proposition 1

Proof of Proposition 1. Let U denote the citizen's expected payo
in any period the incumbent is the leader with the fixed level of advantage p. Then,

$$U = p\left((1-\delta)x + \delta U\right) + (1-p)\left(\gamma + (1-\gamma)\max\left\{(1-\delta)x + \delta U, \delta w(\gamma)\right\}\right).$$

Assume $(1-\delta)x + \delta U \geq \delta w(\gamma)$, then

$$U = p\left((1-\delta)x + \delta U\right) + (1-p)\left(\gamma + (1-\gamma)(1-\delta)x + \delta U\right),$$

so that

$$\begin{aligned} U = \overline{U}_1(p,\gamma,x) &:= \frac{(1-\delta)\left(1-(1-p)\gamma\right)}{1-\delta\left(1-(1-p)\gamma\right)}x + \frac{(1-p)\gamma}{1-\delta\left(1-(1-p)\gamma\right)} \\ &= \frac{(1-\delta)\left(1-(1-p)\gamma\right)}{1-\delta\left(1-(1-p)\gamma\right)}x + 1 - \frac{(1-\delta)\left(1-(1-p)\gamma\right)}{1-\delta\left(1-(1-p)\gamma\right)}. \end{aligned}$$

But

$$\begin{aligned} (1-\delta)x + \delta\overline{U}_1(p,\gamma,x) &= \frac{1-\delta}{1-\delta\left(1-(1-p)\gamma\right)}x + \frac{\delta(1-p)\gamma}{1-\delta\left(1-(1-p)\gamma\right)} \\ &= \delta w(\gamma) + \frac{1-\delta}{1-\delta\left(1-(1-p)\gamma\right)}x + \frac{\delta(1-p)\gamma}{1-\delta\left(1-(1-p)\gamma\right)} - \delta w(\gamma) \\ &= \delta w(\gamma) + \frac{1-\delta}{1-\delta\left(1-(1-p)\gamma\right)}\left(x - p\delta w(\gamma)\right) \\ &\geq \delta w(\gamma) \end{aligned}$$

if and only if $p \leq \frac{x}{\delta w(\gamma)}$. Hence, $U = \overline{U}_1(p,\gamma,x)$ if and only if $p \leq \frac{x}{\delta w(\gamma)}$.

Now assume $(1-\delta)x + \delta U(p) < \delta w(\gamma)$, then

$$U = p\left((1-\delta)x + \delta U\right) + (1-p)w(\gamma),$$

so that

$$\begin{aligned} U = \overline{U}_0(p,\gamma,x) &:= \frac{(1-\delta)p}{1-\delta p}x + \frac{1-p}{1-\delta p}w(\gamma) \\ &= \frac{(1-\delta)p}{1-\delta p}x + \left(1 - \frac{(1-\delta)p}{1-\delta p}\right)w(\gamma). \end{aligned}$$

But

$$(1-\delta)x + \delta\overline{U}_0(p,x,\gamma) = \frac{1-\delta}{1-\delta p}x + \left(1 - \frac{(1-\delta)p}{1-\delta p}\right)\delta w(\gamma)$$
$$= \delta w(\gamma) + \frac{1-\delta}{1-\delta p}\left(x - p\delta w(\gamma)\right)$$
$$< \delta w(\gamma)$$

if and only if $p > \frac{x}{\delta w(\gamma)}$. Hence, $U = \overline{U}_0(p,\gamma,x)$ if and only if $p > \frac{x}{\delta w(\gamma)}$. ■

Lemma 1

Proof of Lemma 1. Let β be given and note that U_β is a fixed point of transformation T_β that maps each bounded function ϕ on $[0,\pi]$ to another bounded function $T_\beta\phi$ such that for each $p \in [0,\pi]$,

$$T_\beta\phi(p) = p\left((1-\delta) + \delta E_\beta(\phi|p)\right) + (1-p)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta E_\beta(\phi|p), \delta w(\gamma)\right\}\right).$$

Clearly, T_β is a contraction mapping, so that U_β is its unique fixed point and for any sequence of bounded functions ϕ^n on $[0,\pi]$ such that $\phi^{n+1} = T_\beta\phi^n$, $\lim_{n\to\infty}\phi^n = U_\beta$.

Consider any function ϕ such that $\overline{U}(\pi,\gamma,x) \le \phi(p) \le \overline{U}(p,\gamma,x)$ for all $p \in [0,\pi]$. First,

$$\begin{aligned}
T_\beta\phi(p) &\ge \pi\left((1-\delta) + \delta E_\beta(\phi|p)\right) + (1-\pi)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta E_\beta(\phi|p), \delta w(\gamma)\right\}\right) \\
&\ge \pi\left((1-\delta) + \delta E_\beta\left(\overline{U}(\pi,\gamma,x)|p\right)\right) \\
&\quad + (1-\pi)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta E_\beta\left(\overline{U}(\pi,\gamma,x)|p\right), \delta w(\gamma)\right\}\right) \\
&= \pi\left((1-\delta) + \delta\overline{U}(\pi,\gamma,x)\right) + (1-\pi)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta\overline{U}(\pi,\gamma,x), \delta w(\gamma)\right\}\right) \\
&= \overline{U}(\pi,\gamma,x)
\end{aligned}$$

holds for all $p \in [0,\pi]$. Hence, $T_\beta\phi \ge \overline{U}(\pi,\gamma,x)$. Second, because $\overline{U}(p,\gamma,x)$ is strictly deceasing in p,

$$\begin{aligned}
T_\beta\phi(p) &\le p\left((1-\delta) + \delta E_\beta\left(\overline{U}(\cdot,\gamma,x)|p\right)\right) \\
&\quad + (1-p)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta E_\beta\left(\overline{U}(\cdot,\gamma,x)|p\right), \delta w(\gamma)\right\}\right) \\
&\le p\left((1-\delta) + \delta E_\beta\left(\overline{U}(p,\gamma,x)|p\right)\right) \\
&\quad + (1-p)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta E_\beta\left(\overline{U}(p,\gamma,x)|p\right), \delta w(\gamma)\right\}\right) \\
&= p\left((1-\delta) + \delta\overline{U}(p,\gamma,x)\right) + (1-p)\left(\gamma + (1-\gamma)\max\left\{(1-\delta) + \delta\overline{U}(p,\gamma,x), \delta w(\gamma)\right\}\right) \\
&= \overline{U}(p,\gamma,x)
\end{aligned}$$

holds for all $p \in [0,\pi]$. Hence, $T_\beta\phi \le \overline{U}(\cdot,\gamma,x)$. Define $\phi^1 = \overline{U}(\cdot,\gamma,x)$ and $\phi^{n+1} = T_\beta\phi^n$.

Because $\overline{U}(\pi, \gamma, x) \leq \phi^1 \leq \overline{U}(\cdot, \gamma, x)$ and because $\overline{U}(\pi, \gamma, x) \leq \phi^n \leq \overline{U}(\cdot, \gamma, x)$ implies that

$$\overline{U}(\pi, \gamma, x) \leq T_\beta \phi^n = \phi^{n+1} \leq \overline{U}(\cdot, \gamma, x),$$

it can be proved inductively that $\overline{U}(\pi, \gamma, x) \leq \phi^n \leq \overline{U}(\cdot, \gamma, x)$ for all n. Therefore,

$$\overline{U}(\pi, \gamma, x) \leq \lim_{n\to\infty} \phi^n = U_\beta \leq \overline{U}(\cdot, \gamma, x).$$

It follows that

$$\overline{U}(\pi, \gamma, x) \leq U_\beta(p) \leq \overline{U}(p, \gamma, x)$$

holds for all $p \in [0, \pi]$. ■

Proposition 2

Proof of Proposition 2.
Backsliding with support. First, suppose $x \geq \pi\delta w(\gamma)$, or equivalently, $\pi \leq \frac{x}{\delta w(\gamma)}$. Then, according to Proposition 1, $(1-\delta)x + \delta\overline{U}(\pi, \gamma, x) \geq \delta w(\gamma)$. In turn, according to Lemma 1,

$$(1-\delta)x + \delta E_\beta(U_\beta|p) \geq (1-\delta)x + \delta E_\beta\left(\overline{U}(\pi, \gamma, x)|p\right) = (1-\delta)x + \delta\overline{U}(\pi, \gamma, x) \geq \delta w(\gamma)$$

holds for all β and all $p \in [0, \pi]$. This implies that in any equilibrium the citizen's strategy is κ^* such that $\kappa^*(p) = 1$ for all $p \in [0, \pi]$.

L_{κ^*} is a fixed point of transformation $\overline{R}$ that maps each bounded function ϕ on $[0, \pi]$ to another bounded function $\overline{R}\phi$ such that for each $p \in [0, \pi]$,

$$\overline{R}\phi(p) = (1 - (1-p)\gamma)\left(1 - \delta + \delta\int_p^\pi \max\{\phi(q), \phi(p)\}\, dF_p(q)\right).$$

Clearly, $\overline{R}$ is a contraction mapping, so that L_{κ^*} is its unique fixed point and for any sequence of bounded functions ϕ^n on $[0, \pi]$ such that $\phi^{n+1} = \overline{R}\phi^n$, $\lim_{n\to\infty}\phi^n = L_{\kappa^*}$.

Consider any function ϕ that is increasing on $[0, \pi]$ and p, p' such that $p < p'$. Then,

$$\begin{aligned}
\overline{R}\phi(p) &= (1 - (1-p)\gamma)\left(1 - \delta + \delta\int_p^\pi \max\{\phi(q), \phi(p)\}\, dF_p(q)\right) \\
&< (1 - (1-p')\gamma)\left(1 - \delta + \delta\int_{p'}^\pi \max\{\phi(q), \phi(p)\}\, dF_{p'}(q)\right) \\
&\leq (1 - (1-p')\gamma)\left(1 - \delta + \delta\int_{p'}^\pi \max\{\phi(q), \phi(p')\}\, dF_{p'}(q)\right) \\
&= \overline{R}\phi(p').
\end{aligned}$$

Define $\phi^1(p) = \frac{(1-\delta)(1-(1-p)\gamma)}{1-\delta(1-(1-p)\gamma)}$ for each $p \in [0, \pi]$, which is strictly increasing in p, and

$\phi^{n+1} = \overline{R}\phi^n$. Because $\phi^1(p) < \phi^1(p')$ and because $\phi^n(p) \leq \phi^n(p')$ implies that

$$\phi^{n+1}(p) = \overline{R}\phi^n(p) < \overline{R}\phi^n(p') = \phi^{n+1}(p')$$

it can be proved inductively that $\phi^n(p) < \phi^n(p')$ for all n. Hence,

$$L_{\kappa^*}(p) = \lim_{n\to\infty} \phi^n(p) \leq \lim_{n\to\infty} \phi^n(p') = L_{\kappa^*}(p')$$

and, in turn,

$$L_{\kappa^*}(p) = \overline{R}L_{\kappa^*}(p) < \overline{R}L_{\kappa^*}(p') = L_{\kappa^*}(p').$$

It follows that $L_{\kappa^*}(p)$ is strictly increasing in p, so that $L_{\kappa^*}(q) > L_{\kappa^*}(p)$ for any p and $q > p$. This implies that given that the citizen has κ^*, it is optimal for the leader to pursue a strategy of backsliding at $p = 0$.

Therefore, when $x \geq \pi\delta w(\gamma)$, there is a unique equilibrium (β^*, κ^*), in which β^* is a strategy of backsliding at $p = 0$ and $\kappa^*(p) = 1$ for all p.

Backsliding against opposition. Now suppose $x < \pi\delta w(\gamma)$, or equivalently, $\frac{x}{\delta w(\gamma)} < \pi$, and let (β^*, κ^*) be any equilibrium. Due to Proposition 1, $(1-\delta)x + \delta\overline{U}\left(\frac{x}{\delta w(\gamma)}, \gamma, x\right) = \delta w(\gamma)$. In turn, according to Lemma 1, for any $p > \frac{x}{\delta w(\gamma)}$

$$\begin{aligned}
(1-\delta)x + E_{\beta^*}(U_{\beta^*}|p) &\leq (1-\delta)x + E_{\beta^*}\left(\overline{U}(\cdot, \gamma, x)|p\right) \\
&\leq (1-\delta)x + E_{\beta^*}\left(\overline{U}(p, \gamma, x)|\frac{x}{\delta w(\gamma)}\right) \\
&= (1-\delta)x + \delta\overline{U}(p, \gamma, x) \\
&< (1-\delta)x + \delta\overline{U}\left(\frac{x}{\delta w(\gamma)}, \gamma, x\right) \\
&= \delta w(\gamma).
\end{aligned}$$

This implies that $\kappa^*(p) = 0$ for all $p > \frac{x}{\delta w(\gamma)}$.

Now let $p > \frac{x}{\delta w(\gamma)}$ be given. Then, restricting on $[p, \pi]$, L_{κ^*} is a fixed point of transformation $\underline{R}$ that maps each bounded function ϕ on $[p, \pi]$ to another bounded function $\underline{R}\phi$ such that for each $p' \in [p, \pi]$,

$$\underline{R}\phi(p') = p'\left(1 - \delta + \delta\int_{p'}^{\pi} \max\left\{\phi(q), \phi(p')\right\} dF_{p'}(q)\right).$$

$\underline{R}$ s a contraction mapping, so that the restriction of L_{κ^*} on $[p, \pi]$ is its unique fixed point and for any sequence of bounded functions ϕ^n on $[p, \pi]$ such that $\phi^{n+1} = \underline{R}\phi^n$, $\lim_{n\to\infty} \phi^n = L_{\kappa^*}$.

Consider any function ϕ that is increasing on $[p, \pi]$ and p', p'' such that $p' < p''$. Then,

$$\begin{aligned}\underline{R}\phi(p') &= p'\left(1-\delta+\delta\int_{p'}^{\pi}\max\{\phi(q),\phi(p')\}\,dF_{p'}(q)\right)\\ &< p''\left(1-\delta+\delta\int_{p''}^{\pi}\max\{\phi(q),\phi(p')\}\,dF_{p''}(q)\right)\\ &\leq p''\left(1-\delta+\delta\int_{p''}^{\pi}\max\{\phi(q),\phi(p'')\}\,dF_{p''}(q)\right)\\ &= \underline{R}\phi(p'').\end{aligned}$$

Define $\phi^1(p') = \frac{(1-\delta)p'}{1-\delta p'}$ for each $p' \in [p, \pi]$, which is strictly increasing in p', and $\phi^{n+1} = \underline{R}\phi^n$. Because $\phi^1(p') < \phi^1(p'')$ and because $\phi^n(p') \leq \phi^n(p'')$ implies that

$$\phi^{n+1}(p') = \underline{R}\phi^n(p') < \underline{R}\phi^n(p'') = \phi^{n+1}(p'')$$

it can be proved inductively that $\phi^n(p') < \phi^n(p'')$ for all n. Hence,

$$L_{\kappa^*}(p') = \lim_{n\to\infty}\phi^n(p') \leq \lim_{n\to\infty}\phi^n(p'') = L_{\kappa^*}(p'')$$

and, in turn,

$$L_{\kappa^*}(p') = \underline{R}L_{\kappa^*}(p') < \underline{R}L_{\kappa^*}(p'') = L_{\kappa^*}(p'').$$

It follows that $L_{\kappa^*}(p')$ is strictly increasing in $p' \in [p, \pi]$, so that $L_{\kappa^*}(q) > L_{\kappa^*}(p')$ for any $p' \geq p$ and $q > p'$. This implies that given that the citizen has κ^*, it is optimal for the leader to pursue a strategy of backsliding at p.

Therefore, when $x < \pi\delta w(\gamma)$, for any equilibrium (β^*, κ^*) and any $p > \frac{x}{\delta w(\gamma)}$, β^* is a strategy of backsliding at p and $\kappa^*(p) = 0$. ■

Proposition 3

Proof of Proposition 3. Let (β^*, κ^*) be an equilibrium that sustains democracy with β^* being a strategy of stopping at a given $p^* < \pi$. Proposition 2 has two implications. First, it must be true that $x < \pi\delta w(\gamma)$, because otherwise β^* has to be a strategy of backsliding at $p = 0$. Second, it must be true that $p^* \leq \frac{x}{\delta w(\gamma)}$, because β^* is a strategy of backsliding at any $p > \frac{x}{\delta w(\gamma)}$. This, according to Lemma 1, implies that $(1-\delta)x + \delta\overline{U}(p^*, \gamma, x) \geq \delta w(\gamma)$.

Consider any function ϕ such that $\phi(p) \geq \overline{U}(p^*, \gamma, x)$ for each $p \in [0, p^*]$. Because $\beta^*(q, p) = 0$ for all $p \leq p^*$ and $q > p^*$,

$$E_{\beta^*}(\phi|p) = \int_p^{p^*}\left(\left(1-\beta^*(q,p)\right)\phi(p) + \beta^*(q,p)\phi(q)\right)dF_p(q) + \left(1-F_p(p^*)\right)\phi(p) \geq \overline{U}(p^*, \gamma, x)$$

holds for all $p \in [0, p^*]$. Hence, for all $p \in [0, p^*]$,

$$\begin{aligned} T_{\beta^*}\phi(p) &\geq p^* \left((1-\delta)x + \delta E_{\beta^*}(\phi|p)\right) + (1-p^*)\left(\gamma + (1-\gamma)\max\left\{(1-\delta)x + \delta E_{\beta^*}(\phi|p), \delta w(\gamma)\right\}\right) \\ &\geq p^* \left((1-\delta)x + \delta \overline{U}(p^*, \gamma, x)\right) \\ &\quad + (1-p^*)\left(\gamma + (1-\gamma)\max\left\{(1-\delta)x + \delta\overline{U}(p^*, \gamma, x), \delta w(\gamma)\right\}\right) \\ &= \overline{U}(p^*, \gamma, x). \end{aligned}$$

Define $\phi^1 = \overline{U}(\cdot, \gamma, x)$ and $\phi^{n+1} = T_{\beta^*}\phi^n$. Because $\phi^1 \geq \overline{U}(p^*, \gamma, x)$ and because $\phi^n \geq \overline{U}(p^*, \gamma, x)$ implies that $\phi^{n+1} = T_{\beta^*}\phi^n$, it can be proved inductively that $\phi^n \geq \overline{U}(p^*, \gamma, x)$ for all n. Hence,

$$U_{\beta^*} = \lim_{n\to\infty} T_{\beta^*}\phi^n \geq \overline{U}(p^*, \gamma, x).$$

Therefore, for all $p \leq p^*$,

$$\begin{aligned} E_{\beta^*}(U_{\beta^*}|p) &= \int_p^{p^*} \left((1-\beta^*(q,p))\, U_{\beta^*}(p) + \beta^*(q,p) U_{\beta^*}(q)\right) dF_p(q) + (1 - F_p(p^*))\, U_{\beta^*}(p) \\ &\geq \overline{U}(p^*, \gamma, x), \end{aligned}$$

so that

$$(1-\delta)x + \delta E_{\beta^*}(U_{\beta^*}|p) \geq (1-\delta)x + \delta\overline{U}(p^*, \gamma, x) \geq \delta w(\gamma),$$

which in turn implies that $\kappa^*(p) = 1$.

Because $\kappa^*(p) = 1$ for all $p \leq p^*$, the restriction of L_{κ^*} on $[0, p^*]$ is the unique fixed point of $\overline{R}$, which is strictly increasing on $[0, p^*]$ (see the proof of Proposition 2). Therefore, for any p, q such that $p < q \leq p^*$, $L_{\kappa^*}(q) > L_{\kappa^*}(p)$, so that $\beta^*(q,p) = 1$. ■

Lemma 2 and Lemma 3

Proof of Lemma 2 and Lemma 3. First, because $\underline{L}$ is the unique fixed point of $\underline{R}$, it must be strictly increasing on $[0, \pi]$ (see the proof of Proposition 2). $L^*(\cdot, \gamma|p^*)$ is a fixed point of transformation $\widehat{R}_{p^*}$ that maps each bounded function ϕ on $[0, p^*]$ to another bounded function $\widehat{R}_{p^*}\phi$ such that for each $p \in [0, p^*]$,

$$\begin{aligned} \widehat{R}_{p^*}\phi(p) &= (1-(1-p)\gamma)\left(1 - \delta + \delta\left(\int_p^{p^*} \phi(q) dF_p(q) + (1 - F_p(p^*))\,\phi(p)\right)\right) \\ &= (1-(1-p)\gamma)\left(1 - \delta + \delta\left(\phi(p) + \int_p^{p^*} (\phi(q) - \phi(p))\, dF_p(q)\right)\right) \end{aligned}$$

Note that $\widehat{R}_{p^*}$ is a contraction mapping, so that $L^*(\cdot, \gamma|p^*)$ is its unique fixed point and for any sequence of bounded functions ϕ^n on $[0, p^*]$ such that $\phi^{n+1} = \widehat{R}_{p^*}\phi^n$, $\lim_{n\to\infty}\phi^n = L^*(\cdot, \gamma|p^*)$.

Consider any function ϕ that is increasing on $[0, p^*]$ and $p, p' \le p^*$ such that $p < p'$. Then,

$$
\begin{aligned}
&\phi(p') + \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_{p'}(q) - \left(\phi(p') + \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_p(q)\right) \\
&= \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_{p'}(q) - \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_p(q) \\
&= (\phi(q) - \phi(p'))\, F_{p'}(q)|_{p'}^{p^*} - \int_{p'}^{p^*} F_{p'}(q) d\phi(q) - \left((\phi(q) - \phi(p'))\, F_p(q)|_{p'}^{p^*} - \int_{p'}^{p^*} F_p(q) d\phi(q)\right) \\
&= (\phi(p^*) - \phi(p'))\, F_{p'}(p^*) - \int_{p'}^{p^*} F_{p'}(q) d\phi(q) - \left((\phi(p^*) - \phi(p'))\, F_p(p^*) - \int_{p'}^{p^*} F_p(q) d\phi(q)\right) \\
&= \int_{p'}^{p^*} (F_{p'}(p^*) - F_{p'}(q))\, d\phi(q) - \int_{p'}^{p^*} (F_p(p^*) - F_p(q))\, d\phi(q) \\
&= \int_{p'}^{p^*} \frac{F(p^*) - F(q)}{1 - F(p')} d\phi(q) - \int_{p'}^{p^*} \frac{F(p^*) - F(q)}{1 - F(p)} d\phi(q) \\
&> 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\phi(p') + \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_p(q) - \left(\phi(p) + \int_{p}^{p^*} (\phi(q) - \phi(p))\, dF_p(q)\right) \\
&= (1 - F_p(p^*) + F_p(p'))\, \phi(p') - (1 - F_p(p^*))\, \phi(p) - \int_p^{p'} \phi(q) dF_p(q) \\
&= (1 - F_p(p^*))\, (\phi(p') - \phi(p)) + \int_p^{p'} (\phi(p') - \phi(p))\, dF_p(q) \\
&> 0.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\widehat{R}_{p^*}\phi(p') &= (1 - (1-\gamma)p') \left(1 - \delta + \delta \left(\phi(p') + \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_{p'}(q)\right)\right) \\
&> (1 - (1-\gamma)p) \left(1 - \delta + \delta \left(\phi(p') + \int_{p'}^{p^*} (\phi(q) - \phi(p'))\, dF_{p'}(q)\right)\right) \\
&> (1 - (1-p)\gamma) \left(1 - \delta + \delta \left(\phi(p) + \int_{p}^{p^*} (\phi(q) - \phi(p))\, dF_{p}(q)\right)\right) \\
&= \widehat{R}_{p^*}\phi(p).
\end{aligned}
$$

Define $\phi^1(p) = \frac{(1-\delta)(1-(1-p)\gamma)}{1-\delta(1-(1-p)\gamma)}$ for each $p \in [0, p^*]$, which is strictly increasing in p, and $\phi^{n+1} = \widehat{R}_{p^*}\phi^n$. Because $\phi^1(p) < \phi^1(p')$ and because $\phi^n(p) \le \phi^n(p')$ implies that

$$
\phi^{n+1}(p) = \widehat{R}_{p^*}\phi^n(p) < \widehat{R}_{p^*}\phi^n(p') = \phi^{n+1}(p')
$$

it can be proved inductively that $\phi^n(p) < \phi^n(p')$ for all n. Hence,

$$L^*(p, \gamma|p^*) = \lim_{n\to\infty} \phi^n(p) \leq \lim_{n\to\infty} \phi^n(p') = L^*(p', \gamma|p^*)$$

and, in turn,

$$L^*(p, \gamma|p^*) = \widehat{R}_{p^*} L^*(p, \gamma|p^*) < \widehat{R}_{p^*} L^*(p', \gamma|p^*) = L^*(p', \gamma|p^*).$$

This implies that $L^*(p, \gamma|p^*)$ is strictly increasing in $p \in [0, \pi]$.

Second, $\underline{U}(\cdot, \gamma, x)$ is a fixed point of transformation $\underline{T}$ that maps each bounded function ϕ on $[0, \pi]$ to another bounded function $\underline{T}\phi$ such that for each $p \in [0, \pi]$,

$$\underline{T}\phi(p) = p\left((1-\delta)x + \delta \int_p^\pi \phi(q) dF_p(q)\right) + (1-p)w(\gamma).$$

Let V be a function such that for each $p \in [0, \pi]$,

$$V(p) = \underline{L}(p)x + (1 - \underline{L}(p))\, w(\gamma),$$

so that

$$\begin{aligned}
\underline{T}V(p) &= p\left((1-\delta)x + \delta\left(\int_p^\pi \underline{L}(q) dF_p(q) x + \left(1 - \int_p^\pi \underline{L}(q) dF_p(q)\right) w(\gamma)\right)\right) + (1-p)w(\gamma) \\
&= p\left(1 - \delta + \delta \int_p^\pi \underline{L}(q) dF_p(q)\right) x + \left(1 - p\left(1 - \delta + \delta \int_p^\pi \underline{L}(q) dF_p(q)\right)\right) w(\gamma) \\
&= \underline{L}(p)x + (1 - \underline{L}(p))\, w(\gamma) \\
&= V(p).
\end{aligned}$$

Hence, V is also a fixed point of $\underline{T}$. Because $\underline{T}$ is a contraction mapping, it has a unique fixed point, which implies that $V = \underline{U}(\cdot, \gamma, x)$.

At last, similarly to the previous arguments, $U^*(\cdot, \gamma, x|p^*)$ is the unique fixed point of contraction mapping $\widehat{T}_{p^*}$ that maps each bounded function ϕ on $[0, p^*]$ to another bounded function $\widehat{T}_{p^*}\phi$ such that for each $p \in [0, p^*]$,

$$\widehat{T}_{p^*}\phi(p) = (1 - (1-p)\gamma)\left((1-\delta)x + \delta\left(\int_p^{p^*} \phi(q) dF_p(q) + (1 - F_p(p^*))\, \phi(p)\right)\right) + (1-p)\gamma.$$

Let V be a function that for each $p \in [0, p^*]$,

$$V(p) = L^*(p, \gamma|p^*)x + 1 - L^*(p, \gamma|p^*)$$

and note that

$$
\begin{aligned}
&\int_p^{p^*} V(q)dF_p(q) + (1-F_p(p^*))\,V(p)\\
&= \left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right)x\\
&\quad + F_p(p^*) - \int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,(1-L^*(p,\gamma|p^*))\\
&= \left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right)x\\
&\quad + 1 - \left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\widehat{T}_{p^*}V(p) &= (1-(1-p)\gamma)\left((1-\delta)x + \delta\left(\int_p^{p^*} V(q)dF_p(q) + (1-F_p(p^*))\,V(p)\right)\right) + (1-p)\gamma\\
&= (1-(1-p)\gamma)\left(1-\delta+\delta\left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right)\right)x\\
&\quad + (1-(1-p)\gamma)\left(\delta-\delta\left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right)\right) + (1-p)\gamma\\
&= (1-(1-p)\gamma)\left(1-\delta+\delta\left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right)\right)x\\
&\quad + 1-(1-(1-p)\gamma)\left(1-\delta+\delta\left(\int_p^{p^*} L^*(q,\gamma|p^*)dF_p(q) + (1-F_p(p^*))\,L^*(p,\gamma|p^*)\right)\right)\\
&= L^*(p,\gamma|p^*)x + 1 - L^*(p,\gamma|p^*)\\
&= V(p)
\end{aligned}
$$

holds for each $p \in [0,p^*]$. Therefore, the uniqueness of $U^*(\cdot,\gamma,x|p^*)$ indicates that $V = U^*(\cdot,\gamma,x|p^*)$. ■

Proposition 4

Proof of Proposition 4.

Defining thresholds. Because $L^*(p,\gamma|p^*)$ is strictly decreasing in p, first

$$
\begin{aligned}
L^*(0,\gamma|p^*) &= (1-\gamma)\left(1-\delta+\delta\left(\int_0^{p^*} L^*(q,\gamma|p^*)dF(q) + (1-F(p^*))\,L^*(0,\gamma|p^*)\right)\right)\\
&\geq (1-\gamma)\,(1-\delta+\delta L^*(0,\gamma|p^*))\\
&\geq \frac{(1-\delta)(1-\gamma)}{1-\delta(1-\gamma)},
\end{aligned}
$$

which implies that

$$L^*(0, 1-\pi|p^*) \geq \frac{(1-\delta)\pi}{1-\delta\pi};$$

second,

$$\begin{aligned} L^*(0,\gamma|p^*) &\leq L^*(p^*,\gamma|p^*) \\ &= (1-(1-p^*)\gamma)\,(1-\delta+\delta L^*(p^*,\gamma|p^*)) \\ &\leq \frac{(1-\delta)\,(1-(1-p^*)\gamma)}{1-\delta\,(1-(1-p^*)\gamma)}, \end{aligned}$$

which implies that

$$L^*\left(0, \frac{1-\pi}{1-p^*}|p^*\right) \leq \frac{(1-\delta)\pi}{1-\delta\pi},$$

where the equality holds if and only if $p^* = 0$. Hence, because $L^*(0,\gamma|p^*)$ is strictly decreasing in γ, there exists a unique $g(p^*) \in \left[1-\pi, \frac{1-\pi}{1-p^*}\right]$ such that

$$L^*\left(0, g(p^*)|p^*\right) = \frac{(1-\delta)\pi}{1-\delta\pi}$$

and $L^*(0,\gamma|p^*) \geq \frac{(1-\delta)\pi}{1-\delta\pi}$ if and only if $\gamma \leq g(p^*)$. By definition, $g(0) = 1-\pi$, and for any $p^* > 0$, because $L^*\left(0, \frac{1-\pi}{1-p^*}|p^*\right) < \frac{(1-\delta)\pi}{1-\delta\pi}$, $g(p^*) < \frac{1-\pi}{1-p^*}$. Moreover, because $L^*(0,\gamma|p^*)$ is strictly increasing in p^*, it must be true that $g(p^*)$ is strictly increasing in p^*.

Clearly from the text, $(1-\delta)x + \int_{p^*}^{\pi} \underline{U}(q,\gamma,x)dF_p(q) \leq \delta w(\gamma)$ if and only if

$$\left(1-\left(1-\int_{p^*}^{\pi}\underline{L}(q)dF_{p^*}(q)\right)\delta\right)x + \left(1-\int_{p^*}^{\pi}\underline{L}(q)dF_{p^*}(q)\right)\delta w(\gamma) \leq \delta w(\gamma),$$

which is equivalent to

$$x \leq h(p^*)\delta w(\gamma),$$

where

$$h(p^*) := \frac{\int_{p^*}^{\pi}\underline{L}(q)dF_{p^*}(q)}{1-\delta\left(1-\int_{p^*}^{\pi}\underline{L}(q)dF_{p^*}(q)\right)}.$$

Because $\underline{L}(p)$ is strictly increasing in p, so is $\int_{p^*}^{\pi}\underline{L}(q)dF_{p^*}(q)$, which implies that $h(p^*)$ is strictly increasing in p^* and is bounded below

$$h(\pi) = \frac{\frac{(1-\delta)\pi}{1-\delta\pi}}{1-\delta\left(1-\frac{(1-\delta)\pi}{1-\delta\pi}\right)} = \pi.$$

At last, because $\underline{L}(p)$ is strictly increasing in p,

$$\begin{aligned}\underline{L}(p) &= p\left(1-\delta+\delta\int_p^\pi \underline{L}(q)dF_p(q)\right)\\ &> p\left(1-\delta+\delta\underline{L}(p)\right)\\ &> \frac{(1-\delta)p}{1-\delta p}\end{aligned}$$

holds for all $p < \pi^*$, which in turn implies that

$$h(p^*) > \frac{\frac{(1-\delta)p^*}{1-\delta p^*}}{1-\delta\left(1-\frac{(1-\delta)p^*}{1-\delta p^*}\right)} = p^*$$

Sufficiency. Let $p^* < \pi$ be given and suppose $\gamma \leq g(p^*)$ and $p^*\delta w(\gamma) \leq x \leq h(p^*)\delta w(\gamma)$. It has been established in the text that (β^*, κ^*) defined as below constitutes an equilibrium:

1. for all $p \leq p^*$, $\kappa^*(p) = 1$ and $\beta^*(q,p) = 1$ for $q \leq p^*$ and $\beta^*(q,p) = 0$ for $q > p^*$;

2. for all $p > p^*$, $\kappa^*(p) = 0$ and $\beta^*(q,p) = 1$ for all $q > p$.

In this equilibrium, β^* is a strategy of stopping at p^*.

Necessity. Let (β^*, κ^*) be an equilibrium in which β^* is a strategy of stopping at p^*. According to Proposition 3, $\beta^*(q,p) = 1$ for all $p < q \leq p^*$ and $\kappa^*(p) = 1$ for all $p \leq p^*$.

First, according to Proposition 2, $\beta^*(q,p) = 1$ for all p, q if $x \geq \pi\delta w(\gamma)$, so that β^* being a stopping strategy necessitates $x < \pi\delta w(\gamma)$. Moreover, because $\kappa^*(p) = 0$ for all $p > \frac{x}{\delta w(\gamma)}$, $\kappa^*(p^*) = 1$ necessitates that $p^* \leq \frac{x}{\delta w(\gamma)}$, or equivalently, $x \geq p^*\delta w(\gamma)$.

Second, for each $p \leq p^*$,

$$L_{\kappa^*}(p) = (1-(1-p)\gamma)\left(1-\delta+\delta\left(\begin{array}{l}\int_p^{p^*} L_{\kappa^*}(q)dF_p(q)\\ +\int_{p^*}^{\pi}\max\left\{L_{\kappa^*}(q), L_{\kappa^*}(p)\right\}dF_p(q)\end{array}\right)\right).$$

Because β^* is a strategy of stopping at p^*, it must be true that $L_{\kappa^*}(p) \geq L_{\kappa^*}(q)$ for all p, q such that $p \leq p^*$ and $q > p^*$. This implies that

$$L_{\kappa^*}(p) = (1-(1-p)\gamma)\left(1-\delta+\delta\left(\int_p^{p^*} L_{\kappa^*}(q)dF_p(q) + (1-F_p(p^*))\,L_{\kappa^*}(p)\right)\right) = \widehat{R}_{p^*}L_{\kappa^*}(p)$$

for all $p \in [0, p^*]$. Due to the uniqueness of the fixed point of $\widehat{R}_{p^*}$, $L_{\kappa^*} = L^*(\cdot, \gamma|p^*)$. Note that because $\pi > \frac{x}{\delta w(\gamma)}$, $\kappa^*(\pi) = 0$, so that $L_{\kappa^*}(\pi) = \pi\left(1-\delta+\delta L_{\kappa^*}(\pi)\right) = \frac{(1-\delta)\pi}{1-\delta\pi}$. Therefore, because $L_{\kappa^*}(0) \geq L_{\kappa^*}(\pi)$,

$$L^*(0,\gamma|p^*) = L_{\kappa^*}(0) \geq L_{\kappa^*}(\pi) = \frac{(1-\delta)\pi}{1-\delta\pi},$$

which in turn implies that $\gamma \leq g(p^*)$.

Third, assume $x > h(p^*)\delta w(\gamma)$, so that $(1-\delta)x + \delta\int_{p^*}^{\pi}\underline{U}(q,\gamma,x)dF_{p^*}(q) > \delta w(\gamma)$. Consider any $p' > p^*$. Restricting on $[p', \pi]$, U_{β^*} is the unique fixed point of T_{β^*}. Because $\underline{U}(p,\gamma,x)$ is strictly decreasing in p,

$$E_{\beta^*}\left(\underline{U}(\cdot,\gamma,x)|p\right) = \int_p^\pi \left((1-\beta(q,p))\,\underline{U}(p,\gamma,x) + \beta(q,p)\underline{U}(q,\gamma,x)\right) dF_p(q)$$
$$\geq \int_p^\pi \underline{U}(q,\gamma,x)dF_p(q),$$

so that

$$\begin{aligned} T_{\beta^*}\underline{U}(p,\gamma,x) &\geq p\left((1-\delta)x + \delta\int_p^\pi \underline{U}(q,\gamma,x)dF_p(q)\right) \\ &\quad + (1-p)\left(\gamma + (1-\gamma)\max\left\{(1-\delta)x + \delta\int_p^\pi \underline{U}(q,\gamma,x)dF_p(q), \delta w(\gamma)\right\}\right) \\ &\geq p\left((1-\delta)x + \delta\int_p^\pi \underline{U}(q,\gamma,x)dF_p(q)\right) + (1-p)w(\gamma) \\ &= \underline{U}(p,\gamma,x) \end{aligned}$$

holds for all $p \in [p', \pi]$. Letting $\phi^1 = \underline{U}(\cdot,\gamma,x)$ and $\phi^{n+1} = T_{\beta^*}\phi^n$, this inductively implies that $\phi^n \geq \underline{U}(\cdot,\gamma,x)$ for all n, so that

$$U_{\beta^*}(p) = \lim_{n\to\infty}\phi^n(p) \geq \underline{U}(p,\gamma,x)$$

for all $p \in [p', \pi]$. It follows that

$$(1-\delta)x + \delta E_{\beta^*}\left(U_{\beta^*}|p'\right) \geq (1-\delta)x + \delta E_{\beta^*}\left(\underline{U}(\cdot,\gamma,x)|p'\right) \geq (1-\delta)x + \delta\int_{p'}^\pi \underline{U}(q,\gamma,x)dF_{p'}(q).$$

Because $(1-\delta)x + \delta\int_{p^*}^\pi \underline{U}(q,\gamma,x)dF_{p^*}(q) > \delta w(\gamma)$, there exists a $p^\dagger > p^*$ sufficiently close to p^*, so that

$$(1-\delta)x + \delta\int_p^\pi \underline{U}(q,\gamma,x)dF_p(q) > \delta w(\gamma)$$

for all $p \in (p^*, p^\dagger]$. Hence, for all $p \in (p^*, p^\dagger]$

$$(1-\delta)x + \delta E_{\beta^*}\left(U_{\beta^*}|p\right) \geq (1-\delta)x + \delta\int_p^\pi \underline{U}(q,\gamma,x)dF_p(q) > \delta w(\gamma),$$

which in turn implies that $\kappa^*(p) = 1$. As a result, $\kappa^*(p) = 1$ for all $p \leq p^\dagger$. As shown in the proof of Proposition 3, this implies that $\beta^*(q,p) = 1$ for all $p < q \leq p^\dagger$. Because $p^\dagger > p^*$, this contradicts the fact that β^* is a strategy of stopping at p^*. Therefore, it must be true that $x \leq h(p^*)\delta w(\gamma)$.

General conditions for sustainability. Democracy is sustainable if and only if there exists a $p^* < \pi$ such that

$$\begin{aligned} &\gamma \leq g(p^*) \\ &p^*\delta w(\gamma) \leq x \leq h(p^*)\delta w(\gamma) \end{aligned}$$

Note that this condition is equivalent to

$$x < \pi\delta w(\gamma)$$
$$\gamma \leq g\left(\frac{x}{\delta w(\gamma)}\right).$$

To establish necessity, first suppose $x \geq \pi\delta w(\gamma)$, then because $h(p^*) < \pi$ for all $p^* < \pi$, $x > h(p^*)\delta w(\gamma)$ for all $p^* < \pi$. Second, suppose $\gamma > g\left(\frac{x}{\delta w(\gamma)}\right)$, then for all $p^* \leq \frac{x}{\delta w(\gamma)}$, $\gamma > g\left(\frac{x}{\delta w(\gamma)}\right) \geq g(p^*)$. This implies that for any p^*, either $p^* > \frac{x}{\delta w(\gamma)}$, so that $x < p^*\delta w(\gamma)$, or $\gamma > g(p^*)$, or both. In any case, the condition for sustainable democracy fails.

To establish sufficiency, let $p^* = \frac{x}{\delta w(\gamma)}$ and note that $p^* < \pi$ because $x < \pi\delta w(\gamma)$. First,

$$g(p^*) = g\left(\frac{x}{\delta w(\gamma)}\right) \geq \gamma.$$

Second,

$$h(p^*)\delta w(\gamma) > p^*\delta w(\gamma) = x.$$

Therefore, democracy is sustained in an equilibrium in which the leader has a strategy of stopping at $p^* = \frac{x}{\delta w(\gamma)}$.

At last, because g is strictly increasing, $\gamma \leq g\left(\frac{x}{\delta w(\gamma)}\right)$ is equivalent to $x \geq g^{-1}(\gamma)\delta w(\gamma)$, so that the condition for sustainable democracy can be rewritten as

$$g^{-1}(\gamma)\delta w(\gamma) \leq x < \pi\delta w(\gamma).$$

The range for x is non-empty if and only if $g^{-1}(\gamma) < \pi$, or equivalently, $\gamma < g(\pi)$. ■

Proposition 5

Proved in the text.