Online Appendices

Meritocracy and Autocratic Power-sharing: A Historical Perspective

Contents

A Proof of the Theorems and Propositions

This section provides the proofs for the theorems and propositions presented in Section 4.

A.1 Proofs for the Meritocratic Institution

Proposition. *4.1.1 (Equilibrium When No Meritocracy Allowed). For every (θ, e^K), there exists a σ_0, such that $P(1+\sigma_0, e^K) - P(1-\sigma_0, e^K) = \frac{1-\theta}{\theta}$.*

- *When $\sigma \geq \sigma_0$, the game has a unique equilibrium where the general coups if and only if he gets a high signal;*
- *When $\sigma < \sigma_0$, the game has a unique equilibrium where the general never coups.*

Proof. Given (β^L, β^H), EU^K strictly decreases in r. Hence, in equilibrium, the ruler will propose r that makes the general indifferent between staging a coup or not at either $e = 1-\sigma$ or $e = 1+\sigma$, but not both. There may be three types of potential equilibrium.

(1) $\beta^H = \beta^L = 0$;

(2) $\beta^H = 1, \beta^L = 0$;

(3) $\beta^H = \beta^L = 1$.

Denote the expected utility of the ruler under the three types of equilibrium as $EU^K_{(i)}, i = 1,2,3$.

(1) $EU^K_{(1)} = w - \theta w P(1+\sigma, e^K)$,
and $r = \theta w P(1+\sigma, e^K)$;

(2) $EU^K_{(2)} = \frac{w}{2} + \frac{\theta}{2} w[P(e^K, 1+\sigma) - P(1-\sigma, e^K)]$,
and $r = \theta w P(1-\sigma, e^K)$;

(3) $EU^K_{(3)} = \frac{\theta}{2} w[P(e^K, 1-\sigma) + P(e^K, 1+\sigma)]$,
and $r = 0$;

We have

$$EU^K_{(1)} - EU^K_{(3)} = \frac{w}{2}[2 - \theta[1 + P(1+\sigma, e^K) + P(e^K, 1-\sigma)]]$$
$$EU^K_{(1)} - EU^K_{(2)} = \frac{w}{2}[(1 - \theta[1 + P(1+\sigma, e^K) - P(1-\sigma, e^K)]]$$
$$EU^K_{(2)} - EU^K_{(3)} = \frac{w}{2}(1-\theta) \geq 0$$

Hence, $EU^K_{(2)} \geq EU^K_{(3)}$ always holds. And $EU^K_{(1)} \geq EU^K_{(2)}$ if and only if $P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{1-\theta}{\theta}$. Because $P(1+\sigma, e^K) - P(1-\sigma, e^K)$ strictly increases in σ, for every $(\theta, e^K) \in (\frac{1}{2}, 1) \times (0, \overline{e^K})$, there exists a σ_0, such that when $\sigma > \sigma_0$, $EU^K_{(1)} < EU^K_{(2)}$, and when $\sigma < \sigma_0$, $EU^K_{(1)} > EU^K_{(2)}$. However, for $\theta \leq \frac{1}{2}$, $\frac{1-\theta}{\theta} \geq 1 \geq P(1+\sigma, e^K) - P(1-\sigma, e^K)$, so $EU^K_{(1)} \geq EU^K_{(2)}$ always holds. □

Theorem. *1 (Equilibrium with Meritocratic Signal). (i) For any (θ, π, e^K) with $\pi < \theta$, there exist $(\underline{\sigma}, \overline{\sigma})$, such that*

$$P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) = \frac{(1-\theta)(1-\pi)}{\pi\theta}$$
$$P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K) = \frac{(1-\theta)\pi}{(1-\pi)\theta}$$

- ***Type I (Coalition of All)*** *When $\sigma < \underline{\sigma}$, there is a unique equilibrium where $r_H = r_L = \theta w P(1+\sigma, e^K)$, and the general never coups.*

- ***Type II (True Meritocracy)*** *When $\underline{\sigma} \leq \sigma < \overline{\sigma}$, there is a unique equilibrium where $r_H = \theta w P(1+\sigma, e^K), r_L = \theta w P(1-\sigma, e^K)$, and the general coups if he is a high type but receives a low signal.*
- ***Type III (Coalition of the Weak)*** *When $\sigma \geq \overline{\sigma}$, there is a unique equilibrium where $r_H = r_L = \theta w P(1-\sigma, e^K)$, and the general coups if he is a high type.*

(ii) For any (θ, π, e^K) with $\pi \geq \theta$, there exist $\tilde{\sigma}$, such that

$$P(1+\tilde{\sigma}, e^K) - P(1-\tilde{\sigma}, e^K) = \frac{(1-\theta)(1-\pi)}{\pi\theta}$$

- *When $\sigma < \tilde{\sigma}$, there is a unique equilibrium where $r_H = r_L = \theta w P(1+\sigma, e^K)$, and the general never coups.*
- *When $\sigma \geq \tilde{\sigma}$, there is a unique equilibrium where $r_H = \theta w P(1+\sigma, e^K), r_L = \theta w P(1-\sigma, e^K)$, and the general coups if he is a high type but receives a low signal.*

Proof. For the same reason as outlined in the previous section, the equilibrium proposal will make the general indifferent to staging a coup or not at some level of (e, s). As long as the signals are partially informative, it is optimal to have $r^H \geq r^L$ for the ruler. Hence, for any e, we have $\beta^{eH} \leq \beta^{eL}$; and for any s, we have $\beta^{Hs} \geq \beta^{Ls}$. Therefore, there may be six types of equilibrium.

(1) $\beta^{HL} = \beta^{HH} = \beta^{LH} = \beta^{LL} = 0$,
$r^H = r^L = \theta w P(1+\sigma, e^K)$;

(2) $\beta^{HL} = 1, \beta^{HH} = \beta^{LH} = \beta^{LL} = 0$,
$r_H = \theta w P(1+\sigma, e^K), r_L = \theta w P(1-\sigma, e^K)$;

(3) $\beta^{HL} = \beta^{HH} = 1, \beta^{LH} = \beta^{LL} = 0$,
$r_H = r_L = \theta w P(1-\sigma, e^K)$;

(4) $\beta^{HL} = \beta^{LL} = 1$, $\beta^{HH} = \beta^{LH} = 0$,
$r_H = \theta w P(1+\sigma, e^K), r_L = 0$;

(5) $\beta^{HL} = \beta^{HH} = \beta^{LL} = 1$, $\beta^{LH} = 0$,
$r_H = \theta w P(1-\sigma, e^K), r_L = 0$;

(6) $\beta^{HL} = \beta^{HH} = \beta^{LH} = \beta^{LL} = 1$,
$r_H = r_L = 0$;

The ruler's expected utility is given by the following:

(1) $EU^K_{(1)} = w - \theta w P(1+\sigma, e^K)$;

(2) $EU^K_{(2)} = \frac{1}{2}[w - \theta w P(1+\sigma, e^K)] + \frac{\pi}{2}[w - \theta w P(1-\sigma, e^K)] + \frac{1-\pi}{2}\theta w P(e^K, 1+\sigma)$;

(3) $EU^K_{(3)} = \frac{1}{2}[w - \theta w P(1-\sigma, e^K)] + \frac{1}{2}\theta w P(e^K, 1+\sigma)$;

(4) $EU^K_{(4)} = \frac{1}{2}[w - \theta w P(1+\sigma, e^K)] + \frac{1-\pi}{2}\theta w P(e^K, 1+\sigma) + \frac{\pi}{2}\theta w P(e^K, 1-\sigma)$;

(5) $EU^K_{(5)} = \frac{1}{2}\theta w P(e^K, 1+\sigma) + \frac{\pi}{2}\theta w P(e^K, 1-\sigma) + \frac{1-\pi}{2}[w - \theta w P(1-\sigma, e^K)]$

(6) $EU^K_{(6)} = \frac{1}{2}\theta w P(e^K, 1+\sigma) + \frac{1}{2}\theta w P(e^K, 1-\sigma)$

We have the following:

$$EU^K_{(3)} - EU^K_{(5)} = \frac{\pi w(1-\theta)}{2} \geq 0$$
$$EU^K_{(2)} - EU^K_{(4)} = \frac{\pi w(1-\theta)}{2} \geq 0$$
$$EU^K_{(3)} - EU^K_{(6)} = \frac{w(1-\theta)}{2} \geq 0$$

Hence, Type (4), (5), and (6) will never be the equilibrium.

For Type (1), (2), (3), we have the following:

$$EU^K_{(1)} - EU^K_{(2)} = \frac{w(1-\theta)(1-\pi)}{2} - \frac{\pi\theta w}{2}[P(1+\sigma, e^K) - P(1-\sigma, e^K)]$$
$$EU^K_{(2)} - EU^K_{(3)} = \frac{w(1-\theta)\pi}{2} - \frac{\theta w(1-\pi)}{2}[P(1+\sigma, e^K) - P(1-\sigma, e^K)]$$

Hence, $EU^K_{(1)} \geq EU^K_{(2)}$ if and only if $P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)(1-\pi)}{\pi\theta}$. And $EU^K_{(2)} \geq EU^K_{(3)}$ if and only if $P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)\pi}{(1-\pi)\theta}$. Since $\pi > \frac{1}{2}$, we have $\frac{(1-\theta)(1-\pi)}{\pi\theta} < \frac{(1-\theta)\pi}{(1-\pi)\theta}$.

Assume $\theta > \frac{1}{2}$. There are two sub-cases:

(i) $\pi < \theta \Leftrightarrow \frac{(1-\theta)\pi}{(1-\pi)\theta} < \lim_{\sigma\to 1} P(1+\sigma, e^K) - P(1-\sigma, e^K)$. Since $P(1+\sigma, e^K) - P(1-\sigma, e^K)$ strictly increases in σ, for any (θ, π, e^K) with $\pi < \theta$ and $e^K < \overline{e^K}$, there exist $(\underline{\sigma}, \overline{\sigma})$, such that

$$P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) = \frac{(1-\theta)(1-\pi)}{\pi\theta}$$
$$P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K) = \frac{(1-\theta)\pi}{(1-\pi)\theta}$$

When $\sigma < \underline{\sigma}$, Type (1) is the unique equilibrium. When $\underline{\sigma} \leq \sigma < \overline{\sigma}$, Type (2) is the unique equilibrium. And When $\sigma \geq \overline{\sigma}$, Type (3) is the unique equilibrium.

(ii) $\pi \geq \theta \Leftrightarrow \frac{(1-\theta)\pi}{(1-\pi)\theta} \geq 1 > \lim_{\sigma\to 1} P(1+\sigma, e^K) - P(1-\sigma, e^K)$. Then $P(1+\sigma, e^K) - P(1-\sigma, e^K) < \frac{(1-\theta)\pi}{(1-\pi)\theta}$ always holds. Since $P(1+\sigma, e^K) - P(1-\sigma, e^K)$ strictly increases in σ, for any (θ, π, e^K) with $\pi \geq \theta$ and $e^K < \overline{e^K}$, there exist $\tilde{\sigma}$, such that

$$P(1+\tilde{\sigma}, e^K) - P(1-\tilde{\sigma}, e^K) = \frac{(1-\theta)(1-\pi)}{\pi\theta}$$

When $\sigma < \tilde{\sigma}$, Type (1) is the unique equilibrium. When $\sigma \geq \tilde{\sigma}$, Type (2) is the unique equilibrium.

□

Proposition. *4.1.3 (Incentive Compatibility of Meritocracy). The conditional redistribution plan $\{r^{L*}, r^{H*}\}$ is ex post incentive compatible if and only if $\pi > \frac{1}{2}$.*

Proof. Ex-post incentive compatibility requires that after receiving the public signal s, the ruler has no incentive to deviate from the redistribution plan $\{r^{L*}, r^{H*}\}$ he pre-committed.

(i) Suppose $\sigma < \underline{\sigma}$, then we have Equilibrium Type (1). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$w - \theta w P(1+\sigma, e^K) \geq \pi\theta w P(e^K, 1+\sigma) + (1-\pi)[w - \theta w P(1-\sigma, e^K)]$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)\pi}{(1-\pi)\theta}$$

When public signal is $s = L$:

$$w - \theta w P(1+\sigma, e^K) \geq (1-\pi)\theta w P(e^K, 1+\sigma) + \pi[w - \theta w P(1-\sigma, e^K)]$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)(1-\pi)}{\pi\theta}$$

Since $\sigma < \underline{\sigma}$, then we have $P(1+\sigma, e^K) - P(1-\sigma, e^K) < \frac{(1-\theta)(1-\pi)}{\pi\theta} \leq \frac{(1-\theta)\pi}{(1-\pi)\theta}$ as required.

(ii) Suppose $\underline{\sigma} \leq \sigma < \overline{\sigma}$, then we have Equilibrium Type (2). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$w - \theta w P(1+\sigma, e^K) \geq \pi\theta w P(e^K, 1+\sigma) + (1-\pi)[w - \theta w P(1-\sigma, e^K)]$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)\pi}{(1-\pi)\theta}$$

When public signal is $s = L$:

$$\pi[w - \theta w P(1-\sigma, e^K)] + (1-\pi)\theta w P(e^K, 1+\sigma) \geq w - \theta w P(1+\sigma, e^K)$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)(1-\pi)}{\pi\theta}$$

Since $\underline{\sigma} \leq \sigma < \overline{\sigma}$, then we have $\frac{(1-\theta)(1-\pi)}{\pi\theta} \leq P(1+\sigma, e^K) - P(1-\sigma, e^K) < \frac{(1-\theta)\pi}{(1-\pi)\theta}$ as required.

(iii) Suppose $\sigma \geq \overline{\sigma}$, then we have Equilibrium Type (3). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$\pi\theta w P(e^K, 1+\sigma) + (1-\pi)[w - \theta w P(1-\sigma, e^K)] \geq w - \theta w P(1+\sigma, e^K)$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)\pi}{(1-\pi)\theta}$$

When public signal is $s = L$:

$$(1-\pi)\theta w P(e^K, 1+\sigma) + \pi[w - \theta w P(1-\sigma, e^K)] \geq w - \theta w P(1+\sigma, e^K)$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)(1-\pi)}{\pi\theta}$$

Since $\sigma \geq \overline{\sigma}$, then we have $P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)\pi}{(1-\pi)\theta} \geq \frac{(1-\theta)(1-\pi)}{\pi\theta}$ as required.

Thus, the ruler does not have an incentive to renege on the redistribution plan after Bayesian updating. Therefore, the conditional redistribution plan $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible. □

Assumption 2. The conflict function has the following functional form: $P(e, e^K) = \frac{e^m}{e^m + (e^K)^m}$, $m > 2$.

Proposition. *4.2.1 (Comparative Statics I). The parameter space which supports true meritocracy expands w.r.t. π.*

Proof. Define $Q(\sigma, e^K) \equiv P(1+\sigma, e^K) - P(1-\sigma, e^K)$. The parameter space supporting true meritocracy is $[\underline{\sigma}, \overline{\sigma}]$.

Further define functions:

$$g\left(\overline{\sigma}, e^K, \theta, \pi\right) \equiv Q(\overline{\sigma}, e^K) - \frac{(1-\theta)\pi}{(1-\pi)\theta}$$
$$h\left(\underline{\sigma}, e^K, \theta, \pi\right) \equiv Q(\underline{\sigma}, e^K) - \frac{(1-\theta)(1-\pi)}{\pi\theta}$$

Fix $\theta > \frac{1}{2}$ and $e^K > 0$. By Implicit Function Theorem, we have

$$\frac{d\overline{\sigma}}{d\pi} = \frac{-g_\pi\left(\overline{\sigma}, \pi; \theta, e^K\right)}{g_{\overline{\sigma}}\left(\overline{\sigma}, \pi; \theta, e^K\right)} \quad \text{and} \quad \frac{d\underline{\sigma}}{d\pi} = \frac{-h_\pi\left(\underline{\sigma}, \pi; \theta, e^K\right)}{h_{\underline{\sigma}}\left(\underline{\sigma}, \pi; \theta, e^K\right)}$$

Notice that

$$Q_\sigma\left(\sigma, e^K\right) = \frac{\partial\left[P\left(1+\sigma, e^K\right) - P\left(1-\sigma, e^K\right)\right]}{\partial\sigma}$$
$$= \frac{\partial P}{\partial e}\left(1+\sigma, e^K\right) + \frac{\partial P}{\partial e}\left(1-\sigma, e^K\right) > 0$$

Thus we have

$$g_{\overline{\sigma}}\left(\overline{\sigma}, \pi; \theta, e^K\right) > 0 \quad \text{and} \quad h_{\underline{\sigma}}\left(\underline{\sigma}, \pi; \theta, e^K\right) > 0$$

Additionally, we get

$$g_\pi\left(\overline{\sigma}, \pi; \theta, e^K\right) = \frac{\partial}{\partial\pi}\left(-\frac{(1-\theta)\pi}{(1-\pi)\theta}\right) < 0 \quad \text{and} \quad h_\pi\left(\underline{\sigma}, \pi; \theta, e^K\right) = \frac{\partial}{\partial\pi}\left(-\frac{(1-\theta)(1-\pi)}{\pi\theta}\right) > 0$$

Therefore we have

$$\frac{d\overline{\sigma}}{d\pi} > 0 \quad \text{and} \quad \frac{d\underline{\sigma}}{d\pi} < 0 \quad \Rightarrow \quad \frac{d(\overline{\sigma} - \underline{\sigma})}{d\pi} > 0$$

Hence, the parameter space that supports true meritocracy (i.e., $\sigma \in [\underline{\sigma}, \overline{\sigma}]$) expands w.r.t. π. ☐

Proposition. *4.2.2 (Comparative Statics II). If the conflict function satisfies Assumption 2, then the parameter space that supports true meritocracy expands w.r.t. e^K when e^K is sufficiently small or sufficiently large.*

Specifically, when e^K satisfies (i) $\frac{e^K}{e} < \sqrt[m]{X_1} < 1$, or (ii) $1 < \frac{e^K}{e} < \frac{\overline{e_1^K}}{e}$, where $X_1 = \frac{2(m-1)(m+1)-\sqrt{3}m\cdot\sqrt{m^2-1}}{(m-1)(m-2)}$ and $\overline{e_1^K}$ satisfies $\lim_{\sigma\to 1} P(1+\sigma, \overline{e_1^K}) - P(1-\sigma, \overline{e_1^K}) = \frac{(1-\theta)\pi}{(1-\pi)\theta}$.

Proof. Fix $\pi > \frac{1}{2}$ and $\theta > \frac{1}{2}$. Recall that we want to show

$$\frac{d(\overline{\sigma} - \underline{\sigma})}{de^K} > 0 \quad \Leftrightarrow \quad \frac{d\overline{\sigma}}{de^K} > \frac{d\underline{\sigma}}{de^K}$$

By Implicit Function Theorem, we can show equivalently that

$$\frac{-g_{e^K}\left(\overline{\sigma}, e^K; \theta, \pi\right)}{g_{\overline{\sigma}}\left(\overline{\sigma}, e^K; \theta, \pi\right)} > \frac{-h_{e^K}\left(\underline{\sigma}, e^K; \theta, \pi\right)}{h_{\underline{\sigma}}\left(\underline{\sigma}, e^K; \theta, \pi\right)}$$

which simplifies to

$$\left.\frac{-Q_{e^K}(\sigma, e^K)}{Q_\sigma(\sigma, e^K)}\right|_{\sigma=\overline{\sigma}} > \left.\frac{-Q_{e^K}(\sigma, e^K)}{Q_\sigma(\sigma, e^K)}\right|_{\sigma=\underline{\sigma}} \tag{1}$$

So far our analysis does not involve the specific functional form: $P\left(e, e^K\right) = \frac{e^m}{e^m + (e^K)^m}$. Now let's take it into account, and first check whether inequality (1) holds for e^K that is sufficiently small.

We have $\frac{\partial^2 P\left(e, e^K\right)}{\partial e \partial e^K} = \frac{m^2 e^{m-1}\left(e^K\right)^{m-1}\left[e^m - \left(e^K\right)^m\right]}{\left[e^m + \left(e^K\right)^m\right]^3} \quad \begin{cases} < 0, & \text{if } e < e^K \\ > 0, & \text{if } e > e^K \end{cases}$

Moreover, we know that

$$\frac{\partial^3 P\left(e, e^K\right)}{\partial e^3}$$
$$= \frac{m\left(e^K\right)^m e^{m-3}\left[\left(e^m\right)^2(m+1)(m+2) + \left(\left(e^K\right)^m\right)^2(m-1)(m-2) + e^m\left(e^K\right)^m(2+2m)(2-2m)\right]}{\left[e^m + \left(e^K\right)^m\right]^4}$$

Set:

$$\left(e^m\right)^2(m+1)(m+2)+\left(\left(e^K\right)^m\right)^2(m-1)(m-2)+e^m\left(e^K\right)^m(2+2m)(2-2m)=0$$
$$\Rightarrow\left(\frac{(e^K)^m}{e^m}\right)^2(m-1)(m-2)+\frac{(e^K)^m}{e^m}\cdot 4(1+m)(1-m)+(m+1)(m+2)=0$$

Let $X=\left(\frac{e^K}{e}\right)^m$, then by the root formula for quadratic equation, we have

$$X_{1,2}=\frac{2(m-1)(m+1)\pm\sqrt{3}m\cdot\sqrt{m^2-1}}{(m-1)(m-2)}$$

For $X\in[X_1,X_2]$, we have $\frac{\partial^3 P(e,e^K)}{\partial e^3}\le 0$. For $X<X_1$ or $X>X_2$, we have $\frac{\partial^3 P(e,e^K)}{\partial e^3}>0$. And notice that $0<X_1<1$ for $m>2$.

(i) For $X=\left(\frac{e^K}{e}\right)^m<X_1$, we have $\frac{\partial^3 P(e,e^K)}{\partial e^3}>0$ and $\frac{\partial^2 P(e,e^K)}{\partial e\partial e^K}>0\Rightarrow\frac{\partial^2 Q(\sigma,e^K)}{\partial\sigma^2}>0$ and $\frac{\partial^2 Q(\sigma,e^K)}{\partial\sigma\partial e^K}>0$.

Thus we have

$$Q_\sigma\left(\sigma,e^K\right)\big|_{\sigma=\overline{\sigma}}>Q_\sigma\left(\sigma,e^K\right)\big|_{\sigma=\underline{\sigma}}>0\quad\text{and}\quad 0>Q_{e^K}\left(\sigma,e^K\right)\big|_{\sigma=\overline{\sigma}}>Q_{e^K}\left(\sigma,e^K\right)\big|_{\sigma=\underline{\sigma}}$$

Therefore the inequality (1) holds, and we've shown that

$$\frac{d(\overline{\sigma}-\underline{\sigma})}{de^K}>0$$

for small e^K s.t. $\frac{e^K}{e}<\sqrt[m]{X_1}<1$.

Now let's check the case (ii) with larger e^K where $1<\frac{e^K}{e}<\frac{\overline{e_1^K}}{e}$.
This case can be further divided into 2 sub cases: (a) $1<\frac{e^K}{e}\le\sqrt[m]{X_2}$; (b) $\sqrt[m]{X_2}<\frac{e^K}{e}<\frac{\overline{e_1^K}}{e}$.
Let's first check sub case (a).

(ii.a) When $1<\frac{e^K}{e}\le\sqrt[m]{X_2}$, we have $\frac{\partial^3 P(e,e^K)}{\partial e^3}<0$ and $\frac{\partial^2 P(e,e^K)}{\partial e\partial e^K}<0\Rightarrow\frac{\partial^2 Q(\sigma,e^K)}{\partial\sigma^2}<0$ and $\frac{\partial^2 Q(\sigma,e^K)}{\partial\sigma\partial e^K}<0$.

Then we have

$$0<Q_\sigma\left(\sigma,e^K\right)\big|_{\sigma=\overline{\sigma}}<Q_\sigma\left(\sigma,e^K\right)\big|_{\sigma=\underline{\sigma}}\quad\text{and}\quad Q_{e^K}\left(\sigma,e^K\right)\big|_{\sigma=\overline{\sigma}}<Q_{e^K}\left(\sigma,e^K\right)\big|_{\sigma=\underline{\sigma}}<0$$

Thus inequality (1) holds, and we've shown that

$$\frac{d(\overline{\sigma}-\underline{\sigma})}{de^K}>0$$

for larger e^K s.t. $1<\frac{e^K}{e}\le\sqrt[m]{X_2}$.

Lastly, let's check sub case (b).

(ii.b) When $\sqrt[m]{X_2}<\frac{e^K}{e}<\frac{\overline{e_1^K}}{e}$, we have $\frac{\partial^3 P(e,e^K)}{\partial e^3}>0$ and $\frac{\partial^2 P(e,e^K)}{\partial e\partial e^K}<0\Rightarrow\frac{\partial^2 Q(\sigma,e^K)}{\partial\sigma^2}>0$ and $\frac{\partial^2 Q(\sigma,e^K)}{\partial\sigma\partial e^K}<0$.

In this parameter space, since $(1\pm\sigma)^m<<(e^K)^m$, then $(e^K)^m\approx(1\pm\sigma)^m+(e^K)^m$.

Thus the following can be simplified as

$$
\begin{aligned}
-Q_{e^K}\left(\sigma, e^K\right) &= -\left[\frac{\partial P}{\partial e^K}\left(1+\sigma, e^K\right) - \frac{\partial P}{\partial e^K}\left(1-\sigma, e^K\right)\right] \\
&\approx m \cdot \frac{(1+\sigma)^m - (1-\sigma)^m}{(e^K)^{m+1}}
\end{aligned}
$$

$$
\begin{aligned}
Q_\sigma\left(\sigma, e^K\right) &= \frac{\partial P}{\partial e}\left(1+\sigma, e^K\right) - \frac{\partial P}{\partial e}\left(1-\sigma, e^K\right) \\
&\approx m \cdot \frac{(1+\sigma)^{m-1} + (1-\sigma)^{m-1}}{(e^K)^m}
\end{aligned}
$$

Define

$$
\begin{aligned}
A(\sigma) &\equiv (1+\sigma)^m - (1-\sigma)^m \\
B(\sigma) &\equiv (1+\sigma)^{m-1} + (1-\sigma)^{m-1} \\
R(\sigma, e^k) &\equiv \frac{-Q_{e^K}\left(\sigma, e^K\right)}{Q_\sigma\left(\sigma, e^K\right)} = \frac{1}{e^K} \cdot \frac{A(\sigma)}{B(\sigma)}
\end{aligned}
$$

Then we have

$$
R_\sigma(\sigma, e^k) = \frac{\left[m((1+\sigma)^{m-1} - (1-\sigma)^{m-1})\right] \cdot B(\sigma) - \left[(m-1)((1+\sigma)^{m-2} - (1-\sigma)^{m-2})\right] \cdot A(\sigma)}{B(\sigma)^2}
$$

Since $\frac{\left[m((1+\sigma)^{m-1}-(1-\sigma)^{m-1})\right]}{(1+\sigma)^m-(1-\sigma)^m} > \frac{\left[(m-1)((1+\sigma)^{m-2}-(1-\sigma)^{m-2})\right]}{(1+\sigma)^{m-1}-(1-\sigma)^{m-1}} > \frac{\left[(m-1)((1+\sigma)^{m-2}-(1-\sigma)^{m-2})\right]}{(1+\sigma)^{m-1}+(1-\sigma)^{m-1}}$, then $R_\sigma(\sigma, e^k) > 0$.

Thus inequality (1) holds and we've shown that

$$
\frac{d(\overline{\sigma} - \underline{\sigma})}{de^K} > 0
$$

for larger e^K s.t. $\sqrt[m]{X_2} < \frac{e^K}{e} < \frac{\overline{e_1^K}}{e}$. $\square$

Proposition. *4.2.3 (Comparative Statics III). If the conflict function satisfies Assumption 2, then:*

- *The parameter space that supports a "Coalition of All" shrinks with respect to e^K when e^K is sufficiently small.*
- *The parameter space that supports a "Coalition of the Weak" first expands as e^K increases from small values and then shrinks as e^K becomes sufficiently large—that is, medium-strong rulers are most likely to form a "Coalition of the Weak."*

The threshold values defining "sufficiently small" and "sufficiently large" e^K are the same as in the preceding proposition.

Proof. The parameter space which supports a "Coalition of All" and a "Coalition of the Weak" are $(0, \underline{\sigma})$ and $(\overline{\sigma}, 1)$, respectively. We know that

$$
\frac{d\sigma}{de^K} = \frac{-Q_{e^K}\left(\sigma, e^K\right)}{Q_\sigma\left(\sigma, e^K\right)}
$$

We have shown above that $Q_\sigma\left(\sigma, e^K\right) > 0$ always holds, thus the sign of this derivative depends on the sign of $Q_{e^K}\left(\sigma, e^K\right)$.

Similar to above, consider the following two cases:

1. When $\left(\frac{e^K}{e}\right)^m < X_1 < 1$, we have $\frac{\partial^2 P(e,e^K)}{\partial e \partial e^K} > 0 \Rightarrow \frac{\partial Q(\sigma,e^K)}{\partial e^K} > 0 \Rightarrow \frac{d\sigma}{de^K} < 0$. Thus, as e^K increases within this range, the parameter spaces $(0,\underline{\sigma})$ and $(\overline{\sigma},1)$ shrink and expand, respectively.
2. When $1 < \frac{e^K}{e} < \frac{\overline{e_1^K}}{e}$, we have $\frac{\partial^2 P(e,e^K)}{\partial e \partial e^K} < 0 \Rightarrow \frac{\partial Q(\sigma,e^K)}{\partial e^K} < 0 \Rightarrow \frac{d\sigma}{de^K} > 0$. Thus, as e^K increases within this range, the parameter space $(\overline{\sigma},1)$ shrinks.

□

A.2 Institutional Evolution under Ignorance about Reinforcement

Assumption 3: $\pi < \theta$, and $\theta w > Z > \frac{w}{2}$.

Let us characterize the transition of the state. State variable σ_t is randomly drawn from a uniform distribution over $[0,\phi] \subset [0,1]$ in each generation. Again, the parameter ϕ captures the amount of uncertainty in the environment. Let R_t^i denotes the per-period payoff for player i, $i \in \{K,A\}$. Variable e_t^K has the following transition rule:

$$e_{t+1}^K = \begin{cases} e_t^K + \mu(\overline{e^K} - e_t^K), & \text{if } R_t^K \geq Z \text{ and } R_t^A < Z \\ e_t^K - \mu e_t^K, & \text{if } R_t^A \geq Z \text{ and } R_t^K < Z \\ e_t^K, & \text{otherwise} \end{cases} \tag{2}$$

where $0 < \mu < 1$ is the rate of adjustments in the ruler's power. Similar to Assumption 1, parameter $\overline{e^K}$ is defined as satisfying $\lim_{\sigma \to \phi} P(1+\sigma,\overline{e^K}) - P(1-\sigma,\overline{e^K}) = \frac{(1-\theta)\pi}{(1-\pi)\theta}$, and Z is the threshold value for power growth. When a player's per-period payoff is greater (or less) than Z, his power will increase (or decrease) in the next generation as a function of μ. Recall that the conflict function $P(e,e^K)$ depends on the relative ratio of the ruler and the general's power, not the absolute level. Hence, the transition rule essentially states that whoever has a per-period payoff greater than the threshold Z will be able to strengthen his power by a parameter of μ. To simplify the calculation, we will treat an increase in the general's power as a reduction in the ruler's. Again, $\overline{e^K}$ ensures that the ruler's power does not grow out of bound, preserving the possibility for all three types of equilibria to emerge. The presence of an upper bound on e^K allows us to investigate the gradual evolution of institutions.

Formally, the state space $Y = \{(e_t^K,\sigma_t)\}$. The action space $A = A^K \times A^A = \{(r_t^H, r_t^L)\} \times \{(\beta_t^{HH}, \beta_t^{HL}, \beta_t^{LL}, \beta_t^{LH})\}$. The reward function $R^K : Y \times A \to \mathbb{R}$ and $R^A : Y \times A \to \mathbb{R}$ are bounded. Let $V^i(y)$ be the value function for player i, $i \in \{K,A\}$.

Lemma (Power Evolution under Meritocracy)**.** *Consider the parameter space where meritocracy endogenously arises. When $\pi < \theta$, there exist two threshold values $(\underline{H}^K, \overline{H}^K)$, such that*

- *When $e_t^K > \overline{H}^K$, the ruler has a probability greater than 1/2 to witness his power grow.*
- *When $e_t^K < \underline{H}^K$, the ruler has a probability greater than 1/2 to witness his power decline.*

Proof. See bellow. □

Lemma (General Power Evolution)**.** *When $\pi < \theta$, there exist two threshold values $(\underline{E}^K, \overline{E}^K)$, such that*

- *When $e_t^K > \overline{E}^K$, the ruler has a probability greater than 1/2 to witness his power grow over time for any value of σ_t.*
- *When $e_t^K < \underline{E}^K$, the ruler has a probability greater than 1/2 to witness his power decline over time for any value of σ_t.*

Proof. When $\pi < \theta$, according to Theorem 1, the static stage game has three types of equilibrium based on the parameters.

First, let's consider the general's coup decision. With ignorance of institutional evolution, the general will ignore the impact of his behavior on the evolution of the institution. He will coup if and only if

$$P(1+e_t,e_t^K)[\theta w+\delta\int V^A(e_t^K,\sigma_t)d\sigma_t]+P(e_t^K,1+e_t)[\delta\int V^A(e_t^K,\sigma_t)d\sigma_t]$$
$$> r_t^{s_t}+\delta\int V^A(e_{t+1}^K,\sigma_t)d\sigma_t \;\Rightarrow\; \theta w P(1+e_t,e_t^K)>r_t^{s_t}$$

The general's coup condition is identical to that in the static stage game. It is straightforward to show that the ruler's optimization will also be the same. Hence, the per-period equilibrium has the same structure as that of the static game.

When $\pi<\theta$, define $(\underline{\sigma_t},\overline{\sigma_t})$ such that

$$P(1+\underline{\sigma_t},e_t^K)-P(1-\underline{\sigma_t},e_t^K)=\frac{(1-\theta)(1-\pi)}{\pi\theta}$$
$$P(1+\overline{\sigma_t},e_t^K)-P(1-\overline{\sigma_t},e_t^K)=\frac{(1-\theta)\pi}{(1-\pi)\theta}$$

There can be three types of equilibrium.

1. When $\sigma_t<\underline{\sigma_t}$, $r_t^H=r_t^L=\theta w P(1+\sigma_t,e_t^K)$, and the general never stages a coup. In this case, there exists a threshold level E_1^K s.t. $\theta w P(1+\sigma_t,E_1^K)=w-Z$, and E_2^K s.t. $\theta w P(1+\sigma_t,E_2^K)=Z$,
   - When $e_t^K>E_1^K$, we have $r^H=r^L<Z$ and $w-r^H=w-r^L>Z$. Hence, the ruler's power is certain to grow over time.
   - When $e_t^K<E_2^K$, we have $r^H=r^L>Z$. Hence, the ruler's power is certain to decline over time.
2. When $\overline{\sigma_t}\ge\sigma_t>\underline{\sigma_t}$, $r_t^H=\theta w P(1+\sigma_t,e_t^K), r_t^L=\theta w P(1-\sigma_t,e_t^K)$, and the general stages a coup when he receives a high type yet a low signal. In this case, there exists a threshold level E_3^K s.t. $\theta w P(1-\sigma_t,E_3^K)=w-Z$, and E_4^K s.t. $\theta w P(1-\sigma_t,E_4^K)=Z$.. We have either $E_1^K>E_2^K>E_3^K>E_4^K$ or $E_1^K>E_3^K\ge E_2^K>E_4^K$.

   Case 2.a: Suppose $E_1^K>E_2^K>E_3^K>E_4^K$.
   - When $e_t^K>E_1^K$, we have $w-r^L>w-r^H>Z$. Hence, the ruler's power has a probability of $\frac{1}{2}[1+\pi+(1-\pi)P(e^K,1+\sigma_t)]>\frac{1}{2}$ to grow over time, and a probability of $\frac{1-\pi}{2}P(1+\sigma_t,e_t^K)<\frac{1}{2}$ to decline.
   - When $E_2^K<e_t^K<E_1^K$, the ruler's power has a probability $\frac{1}{2}[\pi+(1-\pi)P(e_t^K,1+\sigma_t)]<\frac{1}{2}$ to grow over time, and a probability of $\frac{1-\pi}{2}P(1+\sigma_t,e_t^K)<\frac{1}{2}$ to decline.
   - When $E_3^K<e_t^K<E_2^K$, the ruler's power has a probability $\frac{1}{2}[\pi+(1-\pi)P(e_t^K,1+\sigma_t)]<\frac{1}{2}$ to grow over time, and a probability of $\frac{1}{2}[1+(1-\pi)P(1+\sigma_t,e^K)]>\frac{1}{2}$ to decline.
   - When $E_4^K<e_t^K<E_3^K$, the ruler's power has a probability of $\frac{1-\pi}{2}P(e_t^K,1+\sigma_t)<\frac{1}{2}$ to grow, and a probability of $\frac{1}{2}[1+(1-\pi)P(1+\sigma_t,e^K)]>\frac{1}{2}$ to decline over time.
   - When $e_t^K<E_4^K$, the ruler's power has a probability of $\frac{1-\pi}{2}P(e_t^K,1+\sigma_t)<\frac{1}{2}$ to grow, and a probability of $\frac{1}{2}[1+\pi+(1-\pi)P(1+\sigma_t,e_t^K)]>\frac{1}{2}$ to decline over time.

   Case 2.b: Suppose $E_1^K>E_3^K\ge E_2^K>E_4^K$.
   - When $e_t^K>E_1^K$, we have $w-r^L>w-r^H>Z$. Hence, the ruler's power has a probability of $\frac{1}{2}[1+\pi+(1-\pi)P(e^K,1+\sigma_t)]>\frac{1}{2}$ to grow over time, and a probability of $\frac{1-\pi}{2}P(1+\sigma_t,e_t^K)<\frac{1}{2}$ to decline.
   - When $E_3^K<e_t^K<E_1^K$, the ruler's power has a probability $\frac{1}{2}[\pi+(1-\pi)P(e_t^K,1+\sigma_t)]<\frac{1}{2}$ to grow over time, and a probability of $\frac{1}{2}[\pi+(1-\pi)P(1+\sigma_t,e_t^K)]<\frac{1}{2}$ to decline.
   - When $E_2^K<e_t^K<E_3^K$, the ruler's power has a probability $\frac{1-\pi}{2}P(e_t^K,1+\sigma_t)<\frac{1}{2}$ to grow over time, and a probability of $\frac{1-\pi}{2}P(1+\sigma_t,e_t^K)<\frac{1}{2}$ to decline.

- When $E_4^K < e_t^K < E_2^K$, the ruler's power has a probability of $\frac{1-\pi}{2}P(e_t^K, 1+\sigma_t) < \frac{1}{2}$ to grow, and a probability of $\frac{1}{2}[1+(1-\pi)P(1+\sigma_t, e^K)] > \frac{1}{2}$ to decline over time.
- When $e_t^K < E_4^K$, the ruler's power has a probability of $\frac{1-\pi}{2}P(e_t^K, 1+\sigma_t) < \frac{1}{2}$ to grow, and a probability of $\frac{1}{2}[1+\pi+(1-\pi)P(1+\sigma_t, e_t^K)] > \frac{1}{2}$ to decline over time.

3. When $\overline{\sigma_t} \leq \sigma_t$, $r_t^H = r_t^L = \theta w P(1-\sigma_t, e_t^K)$, and the general stages a coup whenever he receives a high type regardless of the signal. In this case,

   - When $E_3^K < e_t^K$, the ruler's power has a probability $\frac{1}{2}[1+P(e_t^K, 1+\sigma_t)] > \frac{1}{2}$ to grow over time, and a probability of $\frac{1}{2}P(1+\sigma_t, e_t^K) < \frac{1}{2}$ to decline over time.
   - When $E_4^K < e_t^K < E_3^K$, the ruler's power has a probability of $\frac{1}{2}P(e_t^K, 1+\sigma_t) < \frac{1}{2}$ to grow, and a probability of $\frac{1}{2}P(1+\sigma_t, e_t^K) < \frac{1}{2}$ to decline over time.
   - When $e_t^K < E_4^K$, the ruler's power has a probability $\frac{1}{2}P(e_t^K, 1+\sigma_t) < \frac{1}{2}$ to grow over time, and a probability $\frac{1}{2}[1+P(1+\sigma_t, e^K)] > \frac{1}{2}$ to decline over time.

Let us define $(\underline{H}^K, \overline{H}^K)$ and $(\underline{E}^K, \overline{E}^K)$ as the following

$$
\begin{aligned}
&\overline{E}^K = \overline{H}^K \ s.t. \\
&\theta w P(1+\sigma_t, E^K) = w - Z, \ and \ P(1+\sigma_t, E^K) - P(1-\sigma_t, E^K) = \frac{(1-\theta)\pi}{(1-\pi)\theta} \\
&\underline{H}^K \ s.t. \\
&\theta w P(1+\sigma_t, H^K) = Z, \ and \ P(1+\sigma_t, H^K) - P(1-\sigma_t, H^K) = \frac{(1-\theta)(1-\pi)}{\pi\theta} \\
&\underline{E}^K \ s.t. \ \theta w P(1-\phi, E^K) = Z
\end{aligned}
$$

When true meritocracy arises, for $e_t^K > \overline{H}^K$, the ruler's power has a probability greater than 1/2 to grow. When $e_t^K < \underline{H}^K$, the ruler's power has a probability greater than 1/2 to decline.

For any type of equilibrium, when $e_t^K > \overline{E}^K$, the ruler's power has a probability greater than 1/2 to grow for all values of σ_t. And When $e_t^K < \underline{E}^K$, the ruler's power has a probability greater than 1/2 to decline for all values of σ_t.

□

When $\pi \geq \theta$, there are only two types of equilibria: a high-payment and no-coup equilibrium or a meritocratic equilibrium. In this case, we can prove the following proposition in a similar fashion.

Proposition. *4.3.1 (Endogenous Institutional Change of Meritocracy). Under Assumptions 2 and 3, suppose* $\underline{H}^K \leq e \cdot \sqrt[m]{X_1}$ *and* $\overline{H}^K < \overline{e^K}$. *Then:*

- *If* $e_t^K > \overline{H}^K$, *there is a probability greater than* 1/2 *that the meritocratic institution is self-reinforcing. That is, it is more likely that the parameter space in* σ_t *which supports true meritocracy will expand over time.*
- *If* $e_t^K < \underline{H}^K$, *there is a probability greater than* 1/2 *that the meritocratic institution is self-undermining. That is, it is more likely that the parameter space in* σ_t *which supports true meritocracy will shrink over time.*

Proof. Omitted. A natural implication of the above propositions. □

B Model Extensions

This section relaxes some of the assumptions of the main model and examines potential extensions.

B.1 Signal Informativeness $\pi_{HH} \neq \pi_{LL}$

In this section, we relax the assumption concerning the information structure. Specifically, we assume $\pi_{HH} \neq \pi_{LL}, 1 > \pi_{HH} > \pi_{LH} > 1 - \pi_{LL} > 0$. Recall that $\theta > \frac{1}{2}$ and $\pi_i < \theta$.

We have the following

$$\begin{aligned}
2EU^K_{(1)} - 2EU^K_{(2)} &= w(1-\theta)(1-\pi_{HH}) - \pi_{LL}\theta w[P(1+\sigma, e^K) - P(1-\sigma, e^K)] \\
2EU^K_{(2)} - 2EU^K_{(3)} &= w(1-\theta)\pi_{HH} - \theta w(1-\pi_{LL})[P(1+\sigma, e^K) - P(1-\sigma, e^K)]
\end{aligned}$$

The equilibrium thresholds are given by the following:

$$\begin{aligned}
P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) &= \frac{(1-\theta)(1-\pi_{HH})}{\theta\pi_{LL}} \\
P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K) &= \frac{(1-\theta)\pi_{HH}}{\theta(1-\pi_{LL})}
\end{aligned}$$

Proposition B.1.1 (Incentive Compatibility of Meritocracy)**.** *The conditional redistribution plan $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible.*

Proof. Ex-post incentive compatibility requires that after receiving the public signal s, the ruler has no incentive to deviate from the redistribution plan $\{r^{L*}, r^{H*}\}$ he pre-committed.

Let $\pi_1 = Prob(e = H|s = H) = \frac{\pi_{HH}}{\pi_{HH}+\pi_{LH}}$, $\pi_2 = Prob(e = L|s = L) = \frac{\pi_{LL}}{\pi_{LL}+\pi_{HL}}$. The condition $\pi_{HH} > \pi_{LH} \Rightarrow \pi_1 > 1/2$, similarly, $\pi_2 > 1/2$.

(i) Suppose $\sigma < \underline{\sigma}$, then we have Equilibrium Type (1). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$\begin{aligned}
w - \theta w P(1+\sigma, e^K) &\geq \pi_1 \theta w P(e^K, 1+\sigma) + (1-\pi_1)[w - \theta w P(1-\sigma, e^K)] \\
\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) &\leq \frac{(1-\theta)\pi_1}{(1-\pi_1)\theta}
\end{aligned}$$

When public signal is $s = L$:

$$\begin{aligned}
w - \theta w P(1+\sigma, e^K) &\geq (1-\pi_2)\theta w P(e^K, 1+\sigma) + \pi_2[w - \theta w P(1-\sigma, e^K)] \\
\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) &\leq \frac{(1-\theta)(1-\pi_2)}{\pi_2\theta}
\end{aligned}$$

The incentive compatibility requirement is equivalent to $P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)(1-\pi_2)}{\pi_2\theta} = \frac{(1-\theta)(\pi_{HL})}{\theta\pi_{LL}}$. On the other hand, since $\sigma < \underline{\sigma}$, the equilibrium threshold condition yields $P(1+\sigma, e^K) - P(1-\sigma, e^K) < \frac{(1-\theta)(1-\pi_{HH})}{\theta\pi_{LL}}$. Therefore, the incentive compatibility is satisfied.

(ii) Suppose $\underline{\sigma} \leq \sigma < \overline{\sigma}$, then we have Equilibrium Type (2). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$\begin{aligned}
w - \theta w P(1+\sigma, e^K) &\geq \pi_1 \theta w P(e^K, 1+\sigma) + (1-\pi_1)[w - \theta w P(1-\sigma, e^K)] \\
\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) &\leq \frac{(1-\theta)\pi_1}{(1-\pi_1)\theta}
\end{aligned}$$

When public signal is $s = L$:

$$\begin{aligned}
\pi_2[w - \theta w P(1-\sigma, e^K)] + (1-\pi_2)\theta w P(e^K, 1+\sigma) &\geq w - \theta w P(1+\sigma, e^K) \\
\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) &\geq \frac{(1-\theta)(1-\pi_2)}{\pi_2\theta}
\end{aligned}$$

The incentive compatibility requirement is equivalent to $\frac{(1-\theta)(1-\pi_2)}{\theta\pi_2} = \frac{(1-\theta)\pi_{HL}}{\theta\pi_{LL}} \leq P(1+\sigma, e^K) - P(1-\sigma, e^K) < \frac{(1-\theta)\pi_1}{\theta(1-\pi_1)} = \frac{(1-\theta)\pi_{HH}}{\theta\pi_{LH}}$.

Since $\underline{\sigma} \leq \sigma < \overline{\sigma}$, the equilibrium threshold conditions yield that $\frac{(1-\theta)(1-\pi_{HH})}{\theta\pi_{LL}} \leq P(1+\sigma, e^K) - P(1-\sigma, e^K) < \frac{(1-\theta)\pi_{HH}}{\theta(1-\pi_{LL})}$ as required. The incentive compatibility requirement is satisfied.

(iii) Suppose $\sigma \geq \overline{\sigma}$, then we have Equilibrium Type (3). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$\begin{aligned}\pi_1\theta w P(e^K, 1+\sigma) + (1-\pi_1)[w - \theta w P(1-\sigma, e^K)] &\geq w - \theta w P(1+\sigma, e^K)\\ \Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) &\geq \frac{(1-\theta)\pi_1}{(1-\pi_1)\theta}\end{aligned}$$

When public signal is $s = L$:

$$\begin{aligned}(1-\pi_2)\theta w P(e^K, 1+\sigma) + \pi_2[w - \theta w P(1-\sigma, e^K)] &\geq w - \theta w P(1+\sigma, e^K)\\ \Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) &\geq \frac{(1-\theta)(1-\pi_2)}{\pi_2\theta}\end{aligned}$$

The incentive compatibility requires $P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)\pi_1}{\theta(1-\pi_1)} = \frac{(1-\theta)\pi_{HH}}{\theta\pi_{LH}}$.

Since $\sigma \geq \overline{\sigma}$, the equilibrium threshold condition yields $P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)\pi_{HH}}{(1-\pi_{LL})\theta}$ as required. The incentive compatibility is satisfied.

Thus, the ruler does not have an incentive to renege on the redistribution plan after Bayesian updating. Therefore, the conditional redistribution plan $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible.

□

B.2 Enhanced Economic Benefit of High Performance

In this section, we let the general's signal directly influence the state of the world. This setting would be closer to a military meritocracy, where the signal not only conveys information about the agent's type but also directly influences the ruler's payoff. The basic finding of this section is that the enhanced economic benefit of competence makes meritocracy more likely to arise—since a larger pie yields more for everyone to share—while the structure of coalitions remains similar to that in the benchmark meritocracy model.

Specifically, the society's total wealth can take two values depending on the general's signal: a high value at w^H or a low value at w^L. $w^H > w^L$. Under the Meritocratic Institution, the ruler's strategy is denoted by $(r^L, r^H) \in [0, w^L] \times [0, w^H]$, representing a signal-conditional redistribution plan towards the general. The general's strategy is given by $\{\beta^{es}\}$, where $\beta^{es} \in [0, 1]$ stands for his probability of initiating a coup after receiving a shock of e and a signal s. For simplicity, we assume that $\pi < \theta$. The case when $\pi \geq \theta$ would be similar to Theorem 2.

Proposition B.2.1 (Equilibrium with Meritocratic Signal and Enhanced Economic Benefit of High Performance)**.** *Assume* $\frac{w^H}{w^L} \geq \frac{P(1-\sigma, w^L)}{P(1-\sigma, w^H)}$*. Then for any* (θ, π, e^K) *with* $\frac{1}{2} < \pi < \theta < 1$*, there exist* $\underline{\sigma}$ *and* $\overline{\sigma}$ *with* $\underline{\sigma} < \overline{\sigma}$ *such that*

$$\begin{aligned}P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) &= \frac{(1-\theta)(1-\pi)}{\pi\theta}\\ P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K) &= \frac{(1-\theta)\pi}{(1-\pi)\theta}\end{aligned}$$

- ***Type I (Coalition of All)*** *When* $\sigma < \underline{\sigma}$*, there is a unique equilibrium where* $r^H = \theta w^H P(1+\sigma, e^K), r^L = \theta w^L P(1+\sigma, e^K)$*, and the general never coups.*

- ***Type II (True Meritocracy)*** *When $\underline{\sigma} \leq \sigma < \overline{\sigma}$, there is a unique equilibrium where $r^H = \theta w^H P(1+\sigma, e^K), r^L = \theta w^L P(1-\sigma, e^K)$, and the general coups if he is a high type but receives a low signal.*
- ***Type III (Coalition of the Weak)*** *When $\sigma > \overline{\sigma}$, there is a unique equilibrium where $r^H = \theta w^H P(1-\sigma, e^K), r^L = \theta w^L P(1-\sigma, e^K)$, and the general coups if he is a high type.*

Proof. For the same reason as outlined in the previous section, there may be six types of equilibrium.

(1) $\beta^{HL} = \beta^{HH} = \beta^{LH} = \beta^{LL} = 0$, $r_1^H = \theta w^H P(1+\sigma, e^K), r_1^L = \theta w^L P(1+\sigma, e^K)$;

(2) $\beta^{HL} = 1, \beta^{HH} = \beta^{LH} = \beta^{LL} = 0$, $r_2^H = \theta w^H P(1+\sigma, e^K), r_2^L = \theta w^L P(1-\sigma, e^K)$;

(3) $\beta^{HL} = \beta^{HH} = 1, \beta^{LH} = \beta^{LL} = 0$, $r_3^H = \theta w^H P(1-\sigma, e^K), r_3^L = \theta w^L P(1-\sigma, e^K)$;

(4) $\beta^{HL} = \beta^{LL} = 1$, $\beta^{HH} = \beta^{LH} = 0$, $r_4^H = \theta w^H P(1+\sigma, e^K), r_4^L = 0$;

(5) $\beta^{HL} = \beta^{HH} = \beta^{LL} = 1$, $\beta^{LH} = 0$, $r_5^H = \theta w^H P(1-\sigma, e^K), r_5^L = 0$;

(6) $\beta^{HL} = \beta^{HH} = \beta^{LH} = \beta^{LL} = 1$, $r_6^H = r_6^L = 0$;

Since $w^H > w^L$, $r^H \geq r^L$ is guaranteed in all cases.
Denote the ruler's expected utility under equilibrium type (i) as $EU^K_{(i)}$.

$$\begin{aligned}
EU^K_{(2)} - EU^K_{(4)} =& \frac{1}{2}\pi w^L(1-\theta) > 0 \\
EU^K_{(3)} - EU^K_{(5)} =& \frac{1}{2}\pi w^L(1-\theta) > 0 \\
EU^K_{(3)} - EU^K_{(6)} =& \frac{1}{2}(1-\pi) w^H(1-\theta) + \frac{1}{2}\pi w^L(1-\theta) > 0
\end{aligned}$$

Thus, Type (4), (5), and (6) will never be the equilibrium.
For Type (1), (2), (3), we have the following :

$$\begin{aligned}
EU^K_{(1)} - EU^K_{(2)} =& \frac{w^L(1-\theta)(1-\pi)}{2} - \frac{\pi\theta w^L}{2}[P(1+\sigma, e^K) - P(1-\sigma, e^K)] \\
EU^K_{(2)} - EU^K_{(3)} =& \frac{w^H(1-\theta)\pi}{2} - \frac{\theta w^H(1-\pi))}{2}[P(1+\sigma, e^K) - P(1-\sigma, e^K)]
\end{aligned}$$

Compare the equations above with those in the proof of Theorem 2, the two settings have similar equilibrium structures. Define cut-off values $\underline{\sigma}$ and $\overline{\sigma}$ similar to above, such that:

$$\begin{aligned}
P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) &= \frac{(1-\theta)(1-\pi)}{\pi\theta} \\
P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K) &= \frac{(1-\theta)\pi}{(1-\pi)\theta}
\end{aligned}$$

Notice that, by assumption $\frac{1}{2} < \pi < \theta < 1$, we have $0 < \frac{(1-\theta)(1-\pi)}{\pi\theta} < \frac{(1-\theta)\pi}{(1-\pi)\theta} < 1$. Thus, we have

$$\begin{aligned}
P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) &= \frac{(1-\theta)(1-\pi)}{\pi\theta} \\
< \frac{(1-\theta)\pi}{(1-\pi)\theta} &= P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K)
\end{aligned}$$

Since $P(1+\sigma, e^K) - P(1-\sigma, e^K)$ is monotonically increasing in σ, then we have $\underline{\sigma} < \overline{\sigma}$.
For the ease of notation, denote:

$$Q = P(1+\sigma, e^K) - P(1-\sigma, e^K)$$

Then there are three possible cases:

(i) $Q < \frac{(1-\theta)(1-\pi)}{\pi\theta} \Leftrightarrow \sigma < \underline{\sigma}$
We have $EU^K_{(1)} > EU^K_{(2)} > EU^K_{(3)}$. Type (1) is the unique equilibrium.

(ii) $\frac{(1-\theta)(1-\pi)}{\pi\theta} < Q < \frac{(1-\theta)\pi}{(1-\pi)\theta} \Leftrightarrow \underline{\sigma} < \sigma < \overline{\sigma}$
We have $EU^K_{(2)} > EU^K_{(1)}$ and $EU^K_{(2)} > EU^K_{(3)}$. Type (2) is the unique equilibrium.

(iii) $\frac{(1-\theta)\pi}{(1-\pi)\theta} < Q \Leftrightarrow \sigma > \overline{\sigma}$
We have $EU^K_{(1)} < EU^K_{(2)} < EU^K_{(3)}$. Type (3) is the unique equilibrium.

□

B.3 Ability and Effort for Signal Generation

In this section, we allow both the general's effort and his ability to influence signal generation. The key insight is that the signal cannot be overly dependent on effort if true meritocracy is to remain incentive-compatible for the king. If a low-type general can successfully mimic a high-type one simply by exerting more effort, then past performance loses much of its informational value in constraining the ruler's opportunistic behavior. This creates a dilemma: while meritocracy is attractive because it incentivizes effort, if performance becomes too effort-driven, the ruler's meritocratic promise may cease to be credible.

Specifically, after nature reveals his ability type e privately, the general makes an effort u, which, together with his ability e, determines the probability that the signal will be high value according to a probability function $l(u+e) \in [0,1]$. Consistent with the main model, we have $l(e^H) = \pi$. After the signal is revealed to all players, the general decides whether to initiate a coup or not. The remainder of the game unfolds similar to that in the main model.

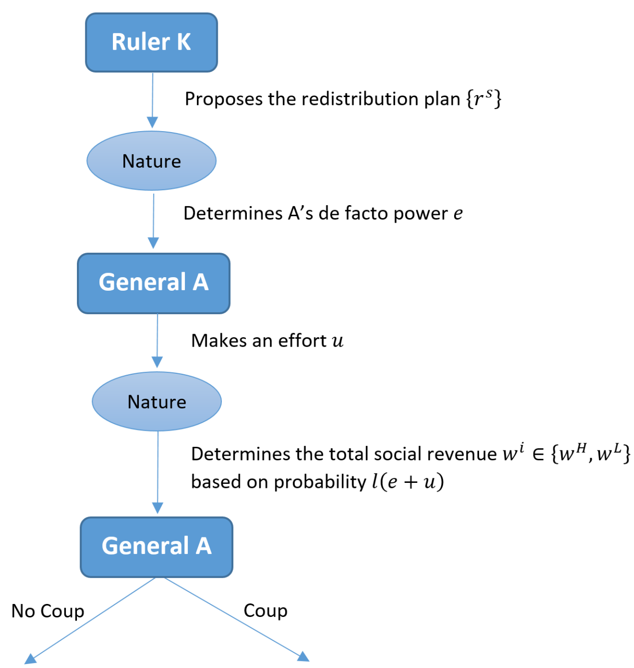


Figure B.1: Timeline under the Meritocratic Institution

Under the Meritocratic Institution, the ruler's strategy is denoted by $\{r^L, r^H\} \in [0,w]$, representing a signal-conditional redistribution plan towards the general. The general's strategy is given by $\{\beta^{es}, u^e\}$, where $\beta^{es} \in [0,1]$ stands for his probability of initiating a coup after receiving an ability shock of e and a signal of s, and u^e is his effort level when he receives an ability shock of e.

Assumption 4. Assume $\pi > 1 - \frac{\theta}{1-\theta}P(2,e^k)$ and $l(e^L + u_2^L) < 1 - \frac{1-\theta}{\theta}\frac{1-\pi}{P(2,e^K)}$.

Proposition B.3.1 (Equilibrium with Meritocratic Signal and Endogenous Effort)**.** *For any (θ, π, e^K) with $e^K < \overline{e^K}$ and $\theta > \frac{1}{2}$, under Assumption 4, there exist $(\underline{\sigma}, \overline{\sigma})$, such that* [1]

- ***Type I (Coalition of All)*** *When $\sigma < \underline{\sigma}$, there is a unique equilibrium where $r^H = r^L = \theta w P(1+\sigma, e^K), u^H = u^L = 0$, and the general never coups.*
- ***Type II (True Meritocracy)*** *When $\underline{\sigma} \leq \sigma < \overline{\sigma}$, there is a unique equilibrium where $r^H = \theta w P(1+\sigma, e^K), r^L = \theta w P(1-\sigma, e^K), u^H = 0, u^L > 0$, and the general coups if he is a high type but receives a low signal.*
- ***Type III (Coalition of the Weak)*** *When $\sigma > \overline{\sigma}$, there is a unique equilibrium where $r^H = r^L = \theta w P(1-\sigma, e^K), u^H = u^L = 0$, and the general coups if he is a high type.*

Proof. For the same reason as outlined in the previous section, the equilibrium proposal will make the general indifferent to staging a coup or not at some level of (e, s). As long as the signals are partially informative, it is optimal to have $r^H \geq r^L$ for the ruler. Hence, for any e, we have $\beta^{eH} \leq \beta^{eL}$; and for any s, we have $\beta^{Hs} \geq \beta^{Ls}$. Therefore, there may be six types of equilibrium.

(1) $\beta^{HL} = \beta^{HH} = \beta^{LH} = \beta^{LL} = 0$, $r_1^H = r_1^L = \theta w P(1+\sigma, e^K), u_1^H = u_1^L = 0$;

(2) $\beta^{HL} = 1, \beta^{HH} = \beta^{LH} = \beta^{LL} = 0$, $r_2^H = \theta w P(1+\sigma, e^K), r_2^L = \theta w P(1-\sigma, e^K)$,
$u_2^H = 0, u_2^L$ *s.t.* $l'(e^L + u_2^L)\theta w[P(1+\sigma, e^K) - P(1-\sigma, e^K)] = c'(u_2^L), u_2^L > 0$;

(3) $\beta^{HL} = \beta^{HH} = 1, \beta^{LH} = \beta^{LL} = 0$, $r_3^H = r_3^L = \theta w P(1-\sigma, e^K)$, $u_3^H = u_3^L = 0$;

(4) $\beta^{HL} = \beta^{LL} = 1$, $\beta^{HH} = \beta^{LH} = 0$, $r_4^H = \theta w P(1+\sigma, e^K), r_4^L = 0$, $u_4^H = 0, u_4^L = u_2^L > 0$;

(5) $\beta^{HL} = \beta^{HH} = \beta^{LL} = 1$, $\beta^{LH} = 0$, $r_5^H = \theta w P(1-\sigma, e^K), r_5^L = 0$, $u_5^H = u_5^L = 0$;

(6) $\beta^{HL} = \beta^{HH} = \beta^{LH} = \beta^{LL} = 1$, $r_6^H = r_6^L = 0$, $u_6^H = u_6^L = 0$;

Additional assumptions to impose:

(i) Convex Cost: $c'(u) > 0$, $c''(u) > 0$

(ii) Concave Effort: $l'(e+u) > 0$, $l''(e+u) < 0$

These assumptions together ensure that SOC< 0, so we get maximizer by solving FOC.

If a general plays Type (2) and (4) strategy:
When $e = L$:

$$\max_{u_2^L} l(e^L + u_2^L) r_2^H + [1 - l(e^L + u_2^L)] r_2^L - c(u_2^L)$$

$$\Rightarrow \; l'(e^L + u_2^L)\theta w[P(1+\sigma, e^K) - P(1-\sigma, e^K)] = c'(u_2^L)$$

When $e = H$:

$$\max_{u_2^H} l(e^H + u_2^H) r_2^H + [1 - l(e^H + u_2^H)]\theta w P(1+\sigma, e^K) - c(u_2^H)$$

$$\Rightarrow FOC \; u_2^H = 0$$

Therefore, we have $u_2^L > 0 = u_1^L = u_3^L, u_1^H = u_2^H = u_3^H = 0$.

[1] The value of $\overline{\sigma}$ may be infinity. When $\overline{\sigma} = \infty$, Type III equilibria do not exist.

Denote the ruler's expected utility under equilibrium type (i) as $EU^K_{(i)}$. We have the following:

$$
\begin{aligned}
EU^K_{(3)} - EU^K_{(5)} &= \frac{(1 - l(e^L + u^L_2))w(1 - \theta)}{2} \geq 0 \\
EU^K_{(2)} - EU^K_{(4)} &= \frac{(1 - l(e^L + u^L_2))w(1 - \theta)}{2} \geq 0 \\
EU^K_{(3)} - EU^K_{(6)} &= \frac{(1 - \theta)}{2} w > 0
\end{aligned}
$$

Hence, Type (4), (5), and (6) will never be the equilibrium.

For types (1), (2), (3), we have the following:

$$
\begin{aligned}
2EU^K_{(1)} - 2EU^K_{(2)} &= w(1 - \theta)(1 - l(e^H)) + w\theta P(1 + \sigma, e^K)[l(e^L + u^L_2) - l(e^L)] \\
&\quad - \theta w[(1 - l(e^L))P(1 + \sigma, e^K) - (1 - l(e^L + u^L_2))P(1 - \sigma, e^K)] \\
&= w(1 - \theta)(1 - l(e^H)) - \theta w[1 - l(e^L + u^L_2)][P(1 + \sigma, e^K) - P(1 - \sigma, e^K)] \\
&\Rightarrow \frac{\partial(2EU^K_{(1)} - 2EU^K_{(2)})}{\partial\sigma} = -\theta w[1 - l(e^L + u^L_2)][P_x(1 + \sigma, e^K) + P_x(1 - \sigma, e^K)] < 0 \\
2EU^K_{(2)} - 2EU^K_{(3)} &= w(1 - \theta)l(e^H) - w[l(e^L + u^L_2) - l(e^L)] \\
&\quad - \theta w[l(e^L + u^L_2)P(1 + \sigma, e^K) - l(e^L)P(1 - \sigma, e^K)] \\
&\Rightarrow \frac{\partial(2EU^K_{(2)} - 2EU^K_{(3)})}{\partial\sigma} = -\theta w[l(e^L + u^L_2)P_x(1 + \sigma, e^K) + l(e^L)P_x(1 - \sigma, e^K)] < 0
\end{aligned}
$$

The following two equations define the threshold value $(\underline{\sigma}, \overline{\sigma})$.

$$
\begin{aligned}
Q1 = P(1 + \underline{\sigma}, e^K) - P(1 - \underline{\sigma}, e^K) &= \frac{(1 - \theta)[1 - l(e^H)]}{\theta[1 - l(e^L + u^L_2)]} \\
Q2 = l(e^L + u^L_2)P(1 + \overline{\sigma}, e^K) - l(e^L)P(1 - \overline{\sigma}, e^K) &= \frac{(1 - \theta)l(e^H)}{\theta} - \frac{l(e^L + u^L_2) - l(e^L)}{\theta}
\end{aligned}
$$

Compare the equations above with those in the proof of Theorem 1. The two settings have similar equilibrium structures. When $\sigma < \underline{\sigma}$, type (1) equilibrium arises. When $\underline{\sigma} < \sigma < \overline{\sigma}$, type (2) equilibrium arises. And type (3) otherwise. The only difference is that the parameter space supporting Type (2) equilibrium shrinks when the general's effort is endogenized.

Lastly, when does a meritocratic equilibrium exist? As long as $Q1 \in [0, P(2, e^K)]$, there exists a $\underline{\sigma}$ that separates a Type I and Type II equilibrium (Note that a Type III equilibrium may not exist under such conditions). Therefore, Assumption 4 is a sufficient and necessary condition to ensure that a Type II meritocratic equilibrium exists. □

Assumption 5. Assume $l(e^L + u^L_2) < \pi = l(e^H)$.

Assumption 4 guarantees that a true meritocracy equilibrium exists. Assumption 5 guarantees that, if a true meritocracy equilibrium exists, it is ex-post incentive compatible for the ruler. If Assumption 5 fails, a Type (2) true meritocracy equilibrium may still exist, but it will not be ex-post incentive-compatible for the King to keep his promise after receiving the signal.

Proposition B.3.2 (Incentive Compatibility of Meritocracy)**.** *Under Assumption 5, if a meritocratic equilibrium exists, it is ex-post incentive compatible.*

Proof. Let $\pi_1 = Prob(e = H|s = H, EqmType13) = \frac{l(e^H)}{l(e^L)+l(e^H)}$, $\pi_2 = Prob(e = L|s = L, EqmType13) = \frac{1-l(e^L)}{2-l(e^L)-l(e^H)}$, $\pi_3 = Prob(e = H|s = H, EqmType2) = \frac{l(e^H)}{l(e^L+u^L_2)+l(e^H)}$, $\pi_4 = Prob(e = L|s = L, EqmType2) = \frac{1-l(e^L+u^L_2)}{2-l(e^L+u^L_2)-l(e^H)}$. The condition $l(e^H) > l(e^L) \Rightarrow \pi_1 > 1/2, \pi_2 > 1/2$. Assumption 5 further ensures that $\pi_3 > 1/2, \pi_4 > 1/2$.

Ex-post incentive compatibility requires that after receiving the public signal s, the ruler has no incentive to deviate from the redistribution plan $\{r^{L*}, r^{H*}\}$ he pre-committed. Let us check whether it holds for all three types of equilibria under meritocratic institution, given $\theta > \frac{1}{2}$ and $\pi < \theta$ (the proof under other parameter spaces would be analogous):

(i) Suppose $\sigma < \underline{\sigma}$, then we have Equilibrium Type (1). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$w - \theta w P(1+\sigma, e^K) \geq \pi_1 \theta w P(e^K, 1+\sigma) + (1-\pi_1) w[1 - \theta P(1-\sigma, e^K)]$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)\pi_1}{(1-\pi_1)\theta}$$

When public signal is $s = L$:

$$w - \theta w P(1+\sigma, e^K) \geq (1-\pi_2)\theta w P(e^K, 1+\sigma) + \pi_2 w[1 - \theta P(1-\sigma, e^K)]$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)(1-\pi_2)}{\pi_2\theta}$$

Since $l(e^H) > l(e^L)$, we have $\frac{1-\pi_2}{\pi_2} = \frac{1-l(e^H)}{1-l(e^L)} \leq \frac{\pi_1}{1-\pi_1} = \frac{l(e^H)}{l(e^L)}$. The incentive compatibility requirement becomes equivilant to $[P(1+\sigma, e^K) - P(1-\sigma, e^K)] \leq \frac{(1-\theta)(1-\pi_2)}{\pi_2\theta} = \frac{1-\theta}{\theta}\frac{1-l(e^H)}{1-l(e^L)}$. Since $\sigma < \underline{\sigma}$, the equilibrium threshold condition yields $[P(1+\sigma, e^K) - P(1-\sigma, e^K)] < \frac{1-\theta}{\theta}\frac{1-l(e^H)}{1-l(e^L+u_2^L)}$. However, as long as $u_2^L > 0$, we have $\frac{1-\theta}{\theta}\frac{1-l(e^H)}{1-l(e^L+u_2^L)} > \frac{1-\theta}{\theta}\frac{1-l(e^H)}{1-l(e^L)}$. Hence, there exists $\sigma < \underline{\sigma}$ s.t. the ruler proposes $r^H = r^L = \theta w P(1+\sigma, e^K)$ ex-ante, but will renege after receiving a low signal.

(ii) Suppose $\underline{\sigma} \leq \sigma < \overline{\sigma}$, then we have Equilibrium Type (2). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if the following constraints are satisfied:

When public signal is $s = H$:

$$w - \theta w P(1+\sigma, e^K) \geq \pi_3 \theta w P(e^K, 1+\sigma) + (1-\pi_3)[w - \theta w P(1-\sigma, e^K)]$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)\pi_3}{(1-\pi_3)\theta}$$

When public signal is $s = L$:

$$\pi_4[w - \theta w P(1-\sigma, e^K)] + (1-\pi_4)\theta w P(e^K, 1+\sigma) \geq w - \theta w P(1+\sigma, e^K)$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)(1-\pi_4)}{\pi_4\theta}$$

The incentive compatibility requirement is equivalent to $\frac{(1-\theta)(1-\pi_4)}{\pi_4\theta} = \frac{(1-\theta)}{\theta}\frac{1-l(e^H)}{1-l(e^L+u_2^L)} \leq P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\theta)\pi_3}{(1-\pi_3)\theta} = \frac{(1-\theta)}{\theta}\frac{l(e^H)}{l(e^L+u_2^L)}$. Since $\underline{\sigma} < \sigma < \overline{\sigma}$, the equilibrium threshold conditions yield $P(1+\sigma, e^K) - P(1-\sigma, e^K) > \frac{(1-\theta)[1-l(e^H)]}{\theta[1-l(e^L+u_2^L)]}$, and $l(e^L+u_2^L)P(1+\sigma, e^K) - l(e^L)P(1-\sigma, e^K) < \frac{(1-\theta)l(e^H)}{\theta} - \frac{l(e^L+u_2^L)-l(e^L)}{\theta}$. Hence, the incentive compatibility is always satisfied.

If Assumption 5 fails but Assumption 4 holds (which is possible when $e^K \to 0^+$), we have $\frac{(1-\theta)(1-\pi_4)}{\pi_4\theta} \geq \frac{(1-\theta)\pi_3}{(1-\pi_3)\theta}$. In this case, a Type (2) true meritocracy equilibrium exists but is not ex-post incentive-compatible for the ruler.

(iii) Suppose $\sigma \geq \overline{\sigma}$, then we have Equilibrium Type (3). $\{r^{L*}, r^{H*}\}$ is ex-post incentive compatible if and only if the following constraints are satisfied:

When public signal is $s = H$:

$$\pi_1 \theta w P(e^K, 1+\sigma) + (1-\pi_1)[w - \theta w P(1-\sigma, e^K)] \geq w - \theta w P(1+\sigma, e^K)$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)\pi_1}{(1-\pi_1)\theta}$$

When public signal is $s = L$:

$$(1-\pi_2)\theta w P(e^K, 1+\sigma) + \pi_2[w - \theta w P(1-\sigma, e^K)] \geq w - \theta w P(1+\sigma, w)$$
$$\Rightarrow P(1+\sigma, e^K) - P(1-\sigma, e^K) \geq \frac{(1-\theta)(1-\pi_2)}{\pi_2 \theta}$$

Since $l(e^H) > l(e^L)$, the incentive compatibility requirement is equivalent to $P(1+\sigma, w) - P(1-\sigma, w) \geq \frac{(1-\theta)\pi_1}{\theta(1-\pi_1)} = \frac{(1-\theta)l(e^H)}{\theta l(e^L)}$. Since $\sigma > \overline{\sigma}$, the equilibrium threshold condition yields $l(e^L + u_2^L)P(1+\sigma, e^K) - l(e^L)P(1-\sigma, e^K) > \frac{w(1-\theta)l(e^H)}{\theta} - \frac{l(e^L+u_2^L)-l(e^L)}{\theta}$. Hence, there exists $\sigma > \overline{\sigma}$, s.t. the ruler proposes $r^H = r^L = \theta w P(1-\sigma, e^K)$ but will renege after receiving a high signal. □

B.4 Loyalty

This section considers the role of loyalty. There can be two orthogonal sources of loyalty, one psychological, ideological, or personality-based (that some people are intrinsically more honest and trustworthy) and the other strategic (that high-ability agents have better outside options and therefore are less "loyal"). The second type of strategic loyalty has already been examined in our main model. In this section, we consider a model extension that incorporates the first type of personality-based loyalty. This modified model yields an equilibrium outcome similar to that of the main model. The key difference is that, in this new setting, both low-ability generals and loyal generals tend to side with the ruler. As a result, the ruler forms "a coalition of the weak and the loyal" when uncertainty is high.

Let $\lambda \in \{F, U\}$ denote the loyalty type of the general, where F represents the faithful general, and U represents the unfaithful general. Let c_λ denote the cost of rebellion for general with loyalty type λ. As rebelling against the ruler is more costly for the loyal general, we have $c_F > c_U$. Without loss of generality, we set $c_U = 0$.

Assumption 6: $c_F > \theta w Q(\sigma, e^K)$

This assumption ensures that the difference in rebellion cost is large enough such that there exists a semi-separating equilibrium in which unfaithful high-ability general is screened out by the ruler's reward plan.

Moreover, assume λ is the general's private information. The ruler's prior belief is that $Prob(\lambda = F) = Prob(\lambda = U) = \frac{1}{2}$. In other words, the ruler is uninformed about the general's loyalty type ex ante and has to make a random guess.

In this modified game, the ruler's strategy is still $(r^L, r^H) \in [0, w]^2$, representing a potentially signal-conditional redistribution plan towards the general. The general's strategy is given by $\{\beta^{es\lambda}\}$, where $\beta^{es\lambda} \in [0, 1]$ stands for his probability of initiating a coup conditional on himself being a general of loyalty type $\lambda \in \{F, U\}$, receiving a shock of $e \in \{L, H\}$ and a signal of $s \in \{L, H\}$.

Proof. The equilibrium proposal will make the general indifferent to staging a coup or not at some level of (e, s, λ). As long as the signals are partially informative, it is optimal to have $r^H \geq r^L$ for the ruler. Hence, for any e, we have $\beta^{eH\lambda} \leq \beta^{eL\lambda}$; and for any s, we have $\beta^{Hs\lambda} \geq \beta^{Ls\lambda}$. Moreover, due to $c_F > c_U$, we have $\beta^{esF} \leq \beta^{esU}$. Therefore, there may be six types of equilibrium.

(1) $\beta^{HLU} = \beta^{HHU} = \beta^{LHU} = \beta^{LLU} = 0 = \beta^{HLF} = \beta^{HHF} = \beta^{LHF} = \beta^{LLF}$, $r^H = r^L = \theta w P(1+\sigma, e^K)$;

(2) $\beta^{HLU} = 1, \beta^{HHU} = \beta^{LHU} = \beta^{LLU} = 0 = \beta^{HLF} = \beta^{HHF} = \beta^{LHF} = \beta^{LLF}$, $r_H = \theta w P(1+\sigma, e^K), r_L = \theta w P(1-\sigma, e^K)$;

(3) $\beta^{HLU} = \beta^{HHU} = 1, \beta^{LHU} = \beta^{LLU} = 0 = \beta^{HLF} = \beta^{HHF} = \beta^{LHF} = \beta^{LLF}$, $r_H = r_L = \theta w P(1-\sigma, e^K)$;

(4) $\beta^{HLU} = \beta^{LLU} = 1 = \beta^{HLF} = \beta^{LLF}$, $\beta^{HHU} = \beta^{LHU} = 0 = \beta^{HHF} = \beta^{LHF}$, $r_H = \theta w P(1+\sigma, e^K), r_L = 0$;

(5) $\beta^{HLU} = \beta^{HHU} = \beta^{LLU} = 1 = \beta^{HLF} = \beta^{HHF} = \beta^{LLF}$, $\beta^{LHU} = 0 = \beta^{LHF}$, $r_H = \theta w P(1-\sigma, e^K), r_L = 0$;

(6) $\beta^{HLU} = \beta^{HHU} = \beta^{LHU} = \beta^{LLU} = 1 = \beta^{HLF} = \beta^{HHF} = \beta^{LHF} = \beta^{LLF}$, $r_H = r_L = 0$.

The ruler's expected utility is given by the following:

(1) $EU^K_{(1)} = w - \theta w P(1+\sigma, e^K)$;

(2) $EU^K_{(2)} = \frac{1}{2}[w - \theta w P(1+\sigma, e^K)] + \frac{1+\pi}{4}[w - \theta w P(1-\sigma, e^K)] + \frac{1-\pi}{4}\theta w P(e^K, 1+\sigma)$;

(3) $EU^K_{(3)} = \frac{3}{4}[w - \theta w P(1-\sigma, e^K)] + \frac{1}{4}\theta w P(e^K, 1+\sigma)$;

(4) $EU^K_{(4)} = \frac{1}{2}[w - \theta w P(1+\sigma, e^K)] + \frac{1-\pi}{2}\theta w P(e^K, 1+\sigma) + \frac{\pi}{2}\theta w P(e^K, 1-\sigma)$;

(5) $EU^K_{(5)} = \frac{1}{2}\theta w P(e^K, 1+\sigma) + \frac{\pi}{2}\theta w P(e^K, 1-\sigma) + \frac{1-\pi}{2}[w - \theta w P(1-\sigma, e^K)]$;

(6) $EU^K_{(6)} = \frac{1}{2}\theta w P(e^K, 1+\sigma) + \frac{1}{2}\theta w P(e^K, 1-\sigma)$.

Similar to the main model, it is trivial to show that Types (4), (5), or (6) will never be the equilibrium. For Type (1), (2), (3), we have the following:

$$EU^K_{(1)} - EU^K_{(2)} = \frac{w(1-\theta)(1-\pi)}{4} - \frac{w\theta(1+\pi)}{4}[P(1+\sigma, e^K) - P(1-\sigma, e^K)]$$
$$EU^K_{(2)} - EU^K_{(3)} = \frac{w(1-\theta)\pi}{4} - \frac{w\theta(2-\pi)}{4}[P(1+\sigma, e^K) - P(1-\sigma, e^K)]$$

Hence, $EU^K_{(1)} \geq EU^K_{(2)}$ if and only if $P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{(1-\pi)(1-\theta)}{(1+\pi)\theta}$. And $EU^K_{(2)} \geq EU^K_{(3)}$ if and only if $P(1+\sigma, e^K) - P(1-\sigma, e^K) \leq \frac{\pi(1-\theta)}{(2-\pi)\theta}$. Since $\pi > \frac{1}{2}$ and $e^K < \overline{e^K_1}$, we have $\frac{(1-\pi)(1-\theta)}{(1+\pi)\theta} < \frac{\pi(1-\theta)}{(2-\pi)\theta} \leq \lim_{\sigma \uparrow 1} P(1+\sigma, e^K) - P(1-\sigma, e^K)$. Assume $\theta > \pi > \frac{1}{2}$. Then there exists $(\underline{\sigma}, \overline{\sigma})$ such that

$$P(1+\underline{\sigma}, e^K) - P(1-\underline{\sigma}, e^K) = \frac{(1-\pi)(1-\theta)}{(1+\pi)\theta}$$
$$P(1+\overline{\sigma}, e^K) - P(1-\overline{\sigma}, e^K) = \frac{\pi(1-\theta)}{(2-\pi)\theta}$$

When $\sigma < \underline{\sigma}$, Type (1) is the unique equilibrium. When $\underline{\sigma} \leq \sigma < \overline{\sigma}$, Type (2) is the unique equilibrium. And when $\sigma \geq \overline{\sigma}$, Type (3) is the unique equilibrium. □

B.5 Endogenizing Signal Choice

So far, we have taken the informativeness of the signal structure π as exogenous. In many historical contexts, however, rulers actively shaped the information environment—for example, by investing in bureaucratic exams, battlefield tests, or other mechanisms of evaluation. To capture this, we introduce an initial stage in which the ruler can choose to improve the informativeness of the signal π above a current level of π_0, incurring a convex cost $C(\pi - \pi_0)$, before announcing his redistribution plan.

The intuition for the results is as follows. A more informative signal structure generally benefits the ruler. However, if the underlying environment is far from the parameter boundary of true meritocracy, then improving the information structure may not alter the equilibrium type that follows. In such cases, the benefit from enhancing π is negligible, and the ruler optimally avoids paying the cost. Thus, investments in informational capacity are endogenous and conditional: rulers rationally choose to build more informative bureaucratic or evaluative institutions only when doing so brings in more benefit than cost. In this case, the information investment decision can be influenced by a series of factors, such as the cost of improvement and other parameters of the society, as indicated by $(\pi_0, e^K, \sigma, \theta)$.

Formally, the timeline of the model becomes the following.

1. The ruler selects $\pi \in [\pi_0, 1)$ at cost $C(\pi - \pi_0)$.
2. The ruler announces a redistribution plan (r_H, r_L).
3. Nature determines the general's type $e \in \{1+\sigma, 1-\sigma\}$.
4. A public signal $s \in \{H, L\}$ is realized with $\Pr(s = H \mid e = 1+\sigma) = \Pr(s = L \mid e = 1-\sigma) = \pi$.
5. The general accepts the contract or stages a coup.

Proposition B.5.1 (Optimal Signal Choice)**.** *Suppose $C(\pi)$ is twice differentiable, increasing, and convex. Then the ruler's optimal choice of π satisfies*

$$\pi^* \in \arg\max_{\pi \in [\pi_0, 1)} \Big\{ EU^K(\pi) - C(\pi - \pi_0) \Big\},$$

where $EU^K(\pi)$ is the ruler's expected equilibrium payoff in Theorem 1 given signal informativeness π.

Corollary B.5.1. *Given $(\pi_0, e^K, \sigma, \theta)$*

(i) If the equilibrium type in a neighborhood of π_0 is Type (2), then the ruler strictly increases π above π_0 if and only if $\frac{w}{2}\Big[1 - \theta + \theta(P(1+\sigma, e_K) - P(1-\sigma, e_K))\Big] > C'(0)$. In this case, the unique interior maximizer π^* solves $C'(\pi^* - \pi_0) = \frac{w}{2}\Big[1 - \theta + \theta(P(1+\sigma, e_K) - P(1-\sigma, e_K))\Big]$. Otherwise, $\pi^* = \pi_0$.

(ii) If the equilibrium type in a neighborhood of π_0 is either Type (1) or Type (3), the ruler adopts the $\pi^* > \pi_0$ specified above if and only if $\frac{w}{2}\Big[1 - \theta + \theta(P(1+\sigma, e_K) - P(1-\sigma, e_K))\Big] > C'(0)$ and $EU_K^{(2)}(\pi^*) - C(\pi^* - \pi_0) - EU_K^{(i)} > 0, \quad i = 1, 3$. Otherwise, the ruler optimally sets $\pi^* = \pi_0$.

Proof. Case (1): π_0 lies in a Type (2) region.

Since $\pi \geq \pi_0$, the subsequent equilibrium will always be Type (2) regardless of information improvement or not. Thus, the ruler's objective at stage 1 is to choose an optimal π to maximize

$$EU^K_{(2)}(\pi) - C(\pi - \pi_0),$$

We have

$$\frac{dEU^K_{(2)}}{d\pi} = \frac{w}{2}\Big[1 - \theta + \theta(P(1+\sigma, e_K) - P(1-\sigma, e_K))\Big] \equiv B > 0.$$

with right derivative at π_0 equal to $B - C'(0)$. Because $P(1+\sigma, e_K) > P(1-\sigma, e_K)$ and $\theta \in (0,1)$, the derivative $\frac{dEU^K_{(2)}}{d\pi}$ is a positive constant w.r.t. π throughout the Type II region.

Therefore, if $C'(0) \geq B$, then any increase in π weakly decreases the objective, so $\pi^* = \pi_0$. If $C'(0) < B$, strict convexity implies the existence of a unique interior maximizer satisfying

$$C'(\pi^* - \pi_0) = B.$$

Here, we assume that $\lim_{x \uparrow 1-\pi_0} C'(x) = +\infty$. Then any finite B can be crossed within the domain of π. Substantively, this assumption implies that the last mile to perfect information is prohibitively expensive, which is quite intuitive.

Case (2): π_0 lies in a Type (1) or Type (3) region.

In the Type I and III regions, $EU^K_{(2)}$ is not influenced by π. Hence, the only reason the ruler wants to improve the signal is that it changes the subsequent equilibrium to a Type (2) one. Hence, for $\pi \geq \pi_0$, the ruler's objective is to maximize

$$EU_K^{(2)} - C(\pi - \pi_0),$$

and maintain

$$EU_K^{(2)} - C(\pi - \pi_0) - EU_K^{(i)} > 0, \quad i = 1, 3.$$

Similar to Case (1), if $C'(0) \geq B$, we have $\pi^* = \pi_0$. If $C'(0) < B$, we have π^* s.t. $C'(\pi^* - \pi_0) = B$ if and only if $EU_K^{(2)}(\pi^*) - C(\pi^* - \pi_0) - EU_K^{(i)} > 0, \quad i = 1, 3$.

□

B.6 Institutional Evolution with Knowledge about Reinforcement

In period t, the power for the ruler is denoted by e_t^K. Similar to the static model, the general has an expected average power of 1, subject to a per-period shock of e_t. Within each generation, the game unfolds under the same timeline as in the benchmark static game.

1. The state of the world is characterized by (e_t^K, σ_t), where σ_t is the level of uncertainty concerning the general's ability. Both (e_t^K, σ_t) are public knowledge among all players;
2. The ruler determines the redistribution plan (r_t^H, r_t^L);
3. Nature determines e_t, and s_t, where $e_t \in \{\sigma_t, -\sigma_t\}$ is the realized type of the general, and $s_t \in \{H, L\}$ is the signal about the general's type. The type e_t is revealed to the general only, while signal s_t is revealed to all players;
4. The general decides to coup or not β_t;
5. If no coup is staged, the payoff $r_t^{s_t}$ is distributed to the general, and $w - r_t^{s_t}$ for the ruler;
6. If a coup is staged, the general succeeds with a probability of $P(1 + e_t, e_t^K)$. The winner of the coup will enjoy θw in the current period, and the loser has zero utility.
7. The game transitions to the next generation with $(e_{t+1}^K, \sigma_{t+1})$.

Now, let us characterize the transition of the state. State variable σ_t is randomly drawn from a uniform distribution over $[0, \phi] \subset [0, 1]$ in each generation. Let R_t^i denotes the per-period payoff for player i, $i \in \{K, A\}$. Variable e_t^K has the following transition rule:

$$e_{t+1}^K = \begin{cases} e_t^K + \mu(\overline{e^K} - e_t^K), & \text{if } R_t^K \geq Z \text{ and } R_t^A < Z \\ e_t^K - \mu e_t^K, & \text{if } R_t^A \geq Z \text{ and } R_t^K < Z \\ e_t^K, & \text{otherwise} \end{cases} \tag{3}$$

where $0 < \mu < 1$ is the rate of adjustments in the ruler's power, $\overline{e^K}$ satisfies $\lim_{\sigma \uparrow \phi} P(1+\sigma, \overline{e^K}) - P(1-\sigma, \overline{e^K}) = \frac{(1-\theta)\pi}{(1-\pi)\theta}$, and Z is the threshold value for power growth. When a player's per-period payoff is greater or less than Z, their power will increase or decrease in the next generation as a function of μ. The introduction of $\overline{e^K}$ ensures that the ruler's power does not grow out of bound, preserving the possibility for all three types of equilibria to emerge, which allows us to investigate the gradual evolution of institutions.

Assumption 7. $Z = \frac{w}{2}$.

This assumption guarantees that one and only one player will be able to strengthen his power in each period. One assumption that we adopt here is different from the norm in the literature. Specifically, the winner of the coup enjoys the total societal payoff only in one period, not forever. Similar to above, we make this assumption so that we can study the gradual evolution of institutions rather than having a society abruptly degenerate into one with only one player. However, we do allow the winner of the coup to enjoy strengthened power in the next generation, capturing the fact that winning a conflict often builds onto one's independent power.

Similar to the static model, we have $e_t \in \{\sigma_t, -\sigma_t\}$ with equal probability, and $Prob(s_t = H | e_t = \sigma_t) = Prob(s_t = L | e_t = -\sigma_t) = \pi > 1/2$. In this game, the ruler's per-period strategy is given by $\{r_t^H, r_t^L\}$, and the general's strategy is given by $\{\beta_t(e_t, s_t)\}$.

Formally, the state space $Y = \{(e_t^K, \sigma_t)\}$. The action space $A = A^K \times A^A = \{(r_t^H, r_t^L)\} \times \{(\beta_t^{HH}, \beta_t^{HL}, \beta_t^{LL}, \beta_t^{LH})\}$. The reward function $R^K : Y \times A \to \mathbb{R}$ and $R^A : Y \times A \to \mathbb{R}$ are bounded. Let $V^i(y) = V^i(e^K, \sigma)$ be the value function for player i, $i \in \{K, A\}$. Let $W^i(e^K) = \int V^i(e^K, \sigma) d\sigma$, $i \in \{K, A\}$, indicates the expected future value aggregated over all possible σ.

The general's Bellman equation

$$TV^A(y) = \max_{\{\beta_t^{es}\}} \int R^A(y_t, r_t^s, \beta_t^{es})d(es) + \delta \int \int_Y V^A(y_{t+1})p(y_{t+1}|y_t, r_t^s, \beta_t^{es})d(es) \tag{4}$$

The ruler's Bellman equation

$$TV^K(y) = \max_{\{r_t^s\}} \int R^K(y_t, r_t^s, \beta_t^{es})d(es) + \delta \int \int_Y V^K(y_{t+1})p(y_{t+1}|y_t, r_t^s, \beta_t^{es})d(es) \tag{5}$$

We focus on Markov Perfect Equilibria, where the per-period strategy depends on the current state of the world and not the entire history of play. In the main text, we solve the game under the assumption that all players are ignorant of institutional evolution. Here, we relax this assumption. We allow players to consider the potential impact of their behavior on future institutions.

To simplify our analysis, we first prove the following Lemmas B.6.1 and B.6.2 in dynamic programming, which are essentially weaker versions of Theorem 9 in Acemoglu's Search Theory Lecture Notes.[2] Note that the notation for the dynamic programming problem is adopted directly from the lecture notes.

Lemma B.6.1. *Let X, Γ, F satisfy Assumption S.2 and S.3.a, and let v be the unique solution to Problem A_2, then v is strictly increasing in x_1.*

- **Problem** A_2:
$$v(x) = \max_{y \in \Gamma(x)} [F(x, y) + \beta v(y)]$$
where $x \in X$ is a vector of state variables, $y \in Y$ is a vector of control variables, $\Gamma : X \rightrightarrows Y$ is a correspondence determining the feasible values of control variables given state variables, F is the payoff function, v is the value function, and $\beta \in (0, 1)$ is the discount factor.

- **Assumption S.2**: X is a compact subset of $\mathbb{R}^k$, Γ is nonempty and compact-valued. Moreover, let $A = \{(x, y) \in X \times Y : y \in \Gamma(x)\}$ and $F : A \to \mathbb{R}$ be bounded and continuous.

- **Assumption S.3.a**: For each y, $F(., y)$ is strictly increasing in at least one of its first k arguments (without loss of generality, let this argument be x_1 where $x = (x_1, x_2, ..., x_k)$). Moreover, Γ is monotone in x_1 in the sense that $x_1 \leq x_1'$ implies $\Gamma(x) \subseteq \Gamma(x')$, where $x' = (x_1', x_2, ..., x_k)$.

Proof. Let $C'(X) \subset C(X)$ be the set of bounded, continuous functions on X, and let $C''(X) \subset C'(X)$ be the set of bounded, continuous functions on X which are strictly increasing in x_1.

Define contraction mapping $T : C(X) \to C(X)$ as following

$$(Tv)(x) = \max_{y \in \Gamma(x)} [F(x, y) + \beta v(y)]$$

We need to show that $T[C'(X)] \subseteq C''(X)$. Given $f \in C'(X)$, we need to show that Tv is strictly increasing in x_1.

By Assumption S.3.a, for any $x_1' > x_1$, we have the following

$$\Gamma((x_1, x_2, \ldots, x_k)) \subseteq \Gamma((x_1', x_2, \ldots, x_k))$$

In addition, for any fixed y,

$$F((x_1', x_2, \ldots, x_k), y) > F((x_1, x_2, \ldots, x_k), y)$$

[2]Acemoglu, D. (2005). 14.461 Advanced Macroeconomics I: Part 1: Search Theory. Retrieved from https://web.mit.edu/14.461/www/part2/lecturenotes1.pdf

Thus we have

$$\begin{aligned}(Tv)(x_1', x_2, \ldots, x_k) &= \max_{y\in\Gamma((x_1', x_2, \ldots, x_k))}[F((x_1', x_2, \ldots, x_k), y) + \beta v(y)] \\ &> \max_{y\in\Gamma((x_1, x_2, \ldots, x_k))}[F((x_1, x_2, \ldots, x_k), y) + \beta v(y)] \\ &= (Tv)(x_1, x_2, \ldots, x_k)\end{aligned}$$

Since $T[C'(X)] \subseteq C''(X)$, then by Theorem 2 in Acemoglu's Search Theory Lecture Notes, the unique fixed point v of T is in $C''(X)$, meaning v is strictly increasing in x_1. ☐

- **Assumption S.3.b**: For each y, $F(., y)$ is strictly decreasing in at least one of its first K arguments (without loss of generality, let this argument be x_1 where $x = (x_1, x_2, ..., x_k)$). Moreover, Γ is monotone in x_1 in the sense that $x_1 \leq x_1'$ implies $\Gamma(x') \subseteq \Gamma(x)$, where $x' = (x_1', x_2, ..., x_k)$.

Lemma B.6.2. *Let X, Γ, F satisfy Assumption S.2 and S.3.b, and let v be the unique solution to Problem A_2, then v is strictly decreasing in x_1.*

Proof. Omitted. Analogous to the proof for Lemma B.6.1. ☐

Proposition B.6.1. *$W^A(e_t^K)$ decreases in e_t^K, while $W^K(e_t^K)$ increases in e_t^K.*

Proof. Since the ruler's current period expected utility is strictly increasing in e_t^K, then $F(\cdot) = \int R^K(y_t, r_t^s, \beta_t^{es})d(es)$ is strictly increasing in e_t^K. In addition, as e_t^K increases, the feasible set of r_t^s (i.e. $\Gamma(\cdot)$) also expands. Thus, by Lemma B.6.1, $V^K(e_t^K, \sigma_t)$ is strictly increasing in e_t^K. Then by the monotonicity of integral, we have $W^K(e_t^K) = \int V^K(e_t^K, \sigma_t)d\sigma_t$ strictly increases in e_t^K.

The proof for $W^A(e_t^K)$ decreases in e_t^K is analogous by applying Lemma B.6.2 to the general's Bellman equation. ☐

Proposition B.6.2 (Equilibrium of the Dynamic Games)**.** *Given (θ, δ, μ), for sufficiently large e_t^K, there exist $(\underline{\sigma}_t, \overline{\sigma}_t)$ such that*

- *When $\sigma_t < \underline{\sigma}_t$, the per-period game has a non-coup equilibrium where the general is always rewarded with high payment;*
- *When $\underline{\sigma}_t < \sigma_t < \overline{\sigma}_t$, the per-period game has a truly meritocratic equilibrium, where a high-signal general is rewarded with high payment and a low-signal one with low payment;*
- *When $\sigma_t > \overline{\sigma}_t$, the per-period game has an equilibrium where the general is always rewarded with low payment. He will coup if and only if he is of a high type. This type of equilibrium may not exist depending on the parameters.*

Proof. First, let's consider the general's coup decision. The general will stage a coup if and only if

$$\begin{aligned}&P(1+e_t, e_t^K)[\theta w + \delta \int V^A\left((1-\mu)e_t^K, \sigma_t\right)d\sigma_t] + P(e_t^K, 1+e_t)[\delta \int V^A\left((1-\mu)e_t^K + \mu\overline{e^K}, \sigma_t\right)d\sigma_t] \\ &> r_t^{s_t} + \delta \int V^A(e_{t+1}^K, \sigma_t)d\sigma_t \cdot (de_{t+1}^K | r_t^{s_t})\end{aligned}$$

Similar to the static game, it is optimal for the ruler to choose a per-period redistribution plan to make either a high-type or a low-type general indifferent.

To simplify notation, define:

$$\begin{aligned}\Delta W^A &= W^A\left((1-\mu)e_t^K\right) - W^A\left((1-\mu)e_t^K + \mu\overline{e^K}\right) \\ \Delta W^K &= W^K\left((1-\mu)e_t^K + \mu\overline{e^K}\right) - W^K\left((1-\mu)e_t^K\right)\end{aligned}$$

By Proposition B.6.1, $\Delta W^A > 0$ and $\Delta W^K > 0$.

Then we have,

$$r_t^{s_t} = \begin{cases} \theta w P(1+e_t, e_t^K) + \delta P(1+e_t, e_t^K)\Delta W^A, \ if \ r_t^{s_t} \geq Z \\ \theta w P(1+e_t, e_t^K) - \delta P(e_t^K, 1+e_t)\Delta W^A, \ if \ r_t^{s_t} < Z \end{cases}$$

Let (η^H, η^L) indicate the value of payment that will make a high-type or low-type general indifferent between staging a coup or not.

We can divide the parameter space into the following areas:

1. When (σ_t, e_t^K) s.t. $\theta P(1+\sigma_t, e_t^K) < \frac{1}{2}$, we have $\eta^H = \theta w P(1+\sigma_t, e_t^K) - \delta P(e_t^K, 1+\sigma_t)\Delta W^A < Z$, and $\eta^L = \theta w P(1-\sigma_t, e_t^K) - \delta P(e_t^K, 1-\sigma_t)\Delta W^A < Z$.
2. When (σ_t, e_t^K) s.t. $\theta P(1+\sigma_t, e_t^K) \geq \frac{1}{2} > \theta P(1-\sigma_t, e_t^K)$, we have $\eta^H = \theta w P(1+\sigma_t, e_t^K) + \delta P(1+\sigma_t, e_t^K)\Delta W^A > Z$,
and $\eta^L = \theta w P(1-\sigma_t, e_t^K) - \delta P(e_t^K, 1-\sigma_t)\Delta W^A < Z$.
3. When (σ_t, e_t^K) s.t. $\theta P(1-\sigma_t, e_t^K) \geq \frac{1}{2}$, we have $\eta^H = \theta w P(1+\sigma_t, e_t^K) + \delta P(1+\sigma_t, e_t^K)\Delta W^A > Z$, and $\eta^L = \theta w P(1-\sigma_t, e_t^K) + \delta P(1-\sigma_t, e_t^K)\Delta W^A > Z$.

Now, let's consider the ruler's decision. Similar to the proof for Theorem 1, we denote the ruler's expected utility under a type i equilibrium as $EU_{(i)}^K$.

1. When (σ_t, e_t^K) s.t. $\theta P(1+\sigma_t, e_t^K) < \frac{1}{2}$,

$$\begin{aligned} 2EU_{(1)}^K - 2EU_{(2)}^K =& (1-\pi)(1-\theta)w - \theta\pi w[P(1+\sigma_t, e_t^K) - P(1-\sigma_t, e_t^K)] \\ &+ \delta(1-\pi)P(1+\sigma_t, e_t^K)\Delta W^K \\ &+ \delta[P(e_t^K, 1+\sigma_t) - \pi P(e_t^K, 1-\sigma_t)]\Delta W^A \\ 2EU_{(2)}^K - 2EU_{(3)}^K =& \pi(1-\theta)w - \theta(1-\pi)w[P(1+\sigma_t, e_t^K) - P(1-\sigma_t, e_t^K)] \\ &+ \delta\pi P(1+\sigma_t, e_t^K)\Delta W^K \\ &+ \delta[P(e_t^K, 1+\sigma_t) - (1-\pi)P(e_t^K, 1-\sigma_t)]\Delta W^A \end{aligned}$$

The ruler's power always increases under peaceful redistribution. This makes a stable equilibrium more attractive to him. It is straightforward to show that equilibrium types (4), (5), and (6) will never be the optimal choice for the ruler.

2. When (σ_t, e_t^K) s.t. $\theta P(1-\sigma_t, e_t^K) < \frac{1}{2} < \theta P(1+\sigma_t, e_t^K)$,

$$\begin{aligned} 2EU_{(1)}^K - 2EU_{(2)}^K =& (1-\pi)(1-\theta)w - \theta\pi w[P(1+\sigma_t, e_t^K) - P(1-\sigma_t, e_t^K)] \\ &+ \delta(1-\pi)P(1+\sigma_t, e_t^K)\Delta W^K \\ &- \delta[\pi P(e_t^K, 1-\sigma_t) + P(1+\sigma_t, e_t^K)]\Delta W^A \\ 2EU_{(2)}^K - 2EU_{(3)}^K =& \pi(1-\theta)w - \theta(1-\pi)w[P(1+\sigma_t, e_t^K) - P(1-\sigma_t, e_t^K)] \\ &- \delta[1-\pi P(1+\sigma_t, e_t^K)]\Delta W^K \\ &- \delta[(1-\pi)P(e_t^K, 1-\sigma_t) + P(1+\sigma_t, e_t^K)]\Delta W^A \end{aligned}$$

Since the ruler's power decreases under a high-payment peaceful redistribution, the continuum play makes a stable equilibrium less attractive for the ruler.

3. When (σ_t, e_t^K) s.t. $\theta P(1-\sigma_t, e_t^K) > \frac{1}{2}$,

$$\begin{aligned}2EU_{(1)}^K - 2EU_{(2)}^K =&(1-\pi)(1-\theta)w - \theta\pi w[P(1+\sigma_t, e_t^K) - P(1-\sigma_t, e_t^K)]\\ &- \delta(1-\pi)P(e_t^K, 1+\sigma_t)\Delta W^K\\ &- \delta[P(1+\sigma_t, e_t^K) - \pi P(1-\sigma_t, e_t^K)]\Delta W^A\\ 2EU_{(2)}^K - 2EU_{(3)}^K =&\pi(1-\theta)w - \theta(1-\pi)w[P(1+\sigma_t, e_t^K) - P(1-\sigma_t, e_t^K)]\\ &- \delta\pi P(e_t^K, 1+\sigma_t)\Delta W^K\\ &- \delta[P(1+\sigma_t, e_t^K) - (1-\pi)P(1-\sigma_t, e_t^K)]\Delta W^A\end{aligned}$$

The ruler's power always decreases under peaceful redistribution. This makes a more stable equilibrium less attractive to him.

Parameter Space 1
Now, let's focus on the case when e_t^K is large enough to fall under parameter space 1. When (σ_t, e_t^K) s.t. $\theta P(1+\sigma_t, e_t^K) < \frac{1}{2}$, since instability is less likely to happen when the power of the general and the ruler is more skewed, we have $\Delta W^A \leq \Delta W^K$.

An assumption to impose: $P\left(e_t^K, 1+\sigma_t\right) - \pi P\left(e_t^K, 1-\sigma_t\right) > 0 \Leftrightarrow \frac{P\left(e_t^K, 1+\sigma_t\right)}{P\left(e_t^K, 1-\sigma_t\right)} > \pi$. This assumption ensures the following two inequalities hold, and it is reasonable to assume: when e_t^k gets large, the difference between $P\left(e_t^K, 1+\sigma_t\right)$ and $P\left(e_t^K, 1-\sigma_t\right)$ is small, and π is typically much lower than 1.

Another assumption to impose: ϕ is large enough s.t. $\phi > \tilde{\sigma}_t \Leftrightarrow P\left(1+\phi, e_t^K\right) - P\left(1-\phi, e_t^K\right) > \frac{(1-\pi)(1-\theta)w+\delta(1-\pi)\Delta W^K}{\pi\theta w+\delta\pi\Delta W^K}$. This assumption posits that there is sufficient amount of uncertainty in the underlying environment for three types of equilibria to emerge.
Then, notice that

$$\begin{aligned}\frac{\partial\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)}{\partial\sigma_t} \leq& -[P_x(1+\sigma_t, e_t^K) + P_x(1-\sigma_t, e_t^K)]\pi\left[\theta w + \delta\Delta W^K\right] < 0\\ \frac{\partial\left(2EU_{(2)}^K - 2EU_{(3)}^K\right)}{\partial\sigma_t} \leq& -[P_x(1+\sigma_t, e_t^K) + P_x(1-\sigma_t, e_t^K)](1-\pi)\left[\theta w + \delta\Delta W^K\right] < 0\end{aligned}$$

Moreover, we have

$$\begin{aligned}\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)\Big|_{\sigma_t=0} =&(1-\pi)(1-\theta)w + \delta(1-\pi)P(1, e_t^K)\Delta W^K\\ &+ \delta(1-\pi)P(e_t^K, 1)\Delta W^A > 0\\ \left(2EU_{(2)}^K - 2EU_{(3)}^K\right)\Big|_{\sigma_t=0} =&\pi(1-\theta)w + \delta\pi P(1, e_t^K)\Delta W^K\\ &+ \delta\pi P(e_t^K, 1)\Delta W^A > 0\end{aligned}$$

However, these conditions do not guarantee the existence of all three types of equilibria. Actually, Type (3) equilibrium does not exist under parameter space 1. To see why, let's check the sign of $2EU_{(2)}^K - 2EU_{(3)}^K$. First, notice that

$$P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e_t^K\right) < 1 < \frac{\pi(1-\theta)w + \delta\pi\Delta W^A}{(1-\pi)\theta w + \delta(1-\pi)\Delta W^A}$$

Thus, we have

$$\begin{aligned}&\pi(1-\theta)w - \theta(1-\pi)w\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e^K\right)\right]\\ &+ \delta\left\{\pi - (1-\pi)\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e_t^K\right)\right]\right\}\Delta W^A > 0\end{aligned}$$

Then by $\Delta W^A \leq \Delta W^K$, we get

$$\begin{aligned}2EU_{(2)}^K - 2EU_{(3)}^K \geq & \pi(1-\theta)w - \theta(1-\pi)w\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e^K\right)\right] \\ & + \delta\left\{\pi - (1-\pi)\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e_t^K\right)\right]\right\}\Delta W^A > 0\end{aligned}$$

So $EU_{(2)}^K > EU_{(3)}^K$ for all σ_t under parameter space 1, rendering Type (3) equilibrium impossible.
Now, let us check the sign of $2EU_{(1)}^K - 2EU_{(2)}^K$.
By $\Delta W^A \leq \Delta W^K$, we have

$$\begin{aligned}2EU_{(1)}^K - 2EU_{(2)}^K \leq & (1-\pi)(1-\theta)w - \theta\pi w\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e_t^K\right)\right] \\ & + \delta\left\{1-\pi-\pi\left[P\left(1+\sigma_t, e^K\right) - P\left(1-\sigma_t, e_t^K\right)\right]\right\}\Delta W^K\end{aligned}$$

Set the right-hand side of the above inequality equal to 0, we get

$$0 < P\left(1+\tilde{\sigma}_t', e_t^K\right) - P\left(1-\tilde{\sigma}_t', e_t^K\right) = \frac{(1-\pi)(1-\theta)w + \delta(1-\pi)\Delta W^K}{\pi\theta w + \delta\pi\Delta W^K} < 1$$

Since $\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)\Big|_{\sigma_t=0} > 0$, $\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)\Big|_{\sigma_t=\tilde{\sigma}_t'} \leq 0$, $\frac{\partial\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)}{\partial\sigma_t} < 0$,
Then $\exists\tilde{\sigma}_t \in (0, \tilde{\sigma}_t'] \subset (0, \phi]$ s.t. $\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)\Big|_{\sigma_t=\tilde{\sigma}_t} = 0$ and when $\sigma_t < \tilde{\sigma}_t$, $EU_{(1)}^K > EU_{(2)}^K$; when $\sigma_t > \tilde{\sigma}_t$, $EU_{(1)}^K < EU_{(2)}^K$.

Hence, under parameter space 1, given any e_t^K, there exists $\tilde{\sigma}_t$ such that

- When $\sigma_t < \tilde{\sigma}_t$, there is a unique equilibrium where $r_t^H = r_t^L = \eta^H$, and the general never coups.
- When $\sigma_t \geq \tilde{\sigma}_t$, there is a unique equilibrium where $r_t^H = \eta^H, r_t^L = \eta^L$, and the general coups if he is a high type but receives a low signal.

Parameter Space 3
Next, let us focus on the case when e_t^K is small enough to fall under parameter space 3. When (σ_t, e_t^K) s.t. $\theta P(1-\sigma_t, e_t^K) > \frac{1}{2}$, since instability is more likely to happen when the power of the general and the ruler is less skewed, we have $\Delta W^A \geq \Delta W^K$.

Again, we assume that ϕ is large enough s.t. $\phi > \overline{\sigma}_t \Leftrightarrow P\left(1+\phi, e_t^K\right) - P\left(1-\phi, e_t^K\right) > \frac{\pi(1-\theta)w - \delta\pi\Delta W^K}{(1-\pi)\theta w + \delta(1-\pi)\Delta W^K}$)
Notice that

$$\begin{aligned}\frac{\partial\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)}{\partial\sigma_t} & \leq -\left[P_x(1+\sigma_t, e_t^K) + P_x(1-\sigma_t, e_t^K)\right]\pi\left[\theta w + \delta\Delta W^K\right] < 0 \\ \frac{\partial\left(2EU_{(2)}^K - 2EU_{(3)}^K\right)}{\partial\sigma_t} & \leq -\left[P_x(1+\sigma_t, e_t^K) + P_x(1-\sigma_t, e_t^K)\right](1-\pi)\left[\theta w + \delta\Delta W^K\right] < 0\end{aligned}$$

$$\begin{aligned}\left(2EU_{(1)}^K - 2EU_{(2)}^K\right)\Big|_{\sigma_t=0} & \geqslant (1-\pi)\left[(1-\theta)w - \delta\Delta W^A\right] \\ \left(2EU_{(2)}^K - 2EU_{(3)}^K\right)\Big|_{\sigma_t=0} & \geqslant \pi\left[(1-\theta)w - \delta\Delta W^A\right]\end{aligned}$$

To retain the same equilibrium structure as above, we impose an additional assumption: $(1-\theta)w - \delta\Delta W^A > 0 \Leftrightarrow \delta < \frac{(1-\theta)w}{\Delta W^A}$, which ensures $EU_1^K > EU_2^K > EU_3^K$ when $\sigma_t = 0$.
Next, let us check the sign of $2EU_{(1)}^K - 2EU_{(2)}^K$.

$$\begin{aligned}2EU_{(1)}^K - 2EU_{(2)}^K \leqslant & (1-\pi)(1-\theta)w - \theta\pi w\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e_t^K\right)\right] \\ & - \delta\left\{1-\pi+\pi\left[P\left(1+\sigma_t, e_t^K\right) - P\left(1-\sigma_t, e_t^K\right)\right]\right\}\Delta W^K\end{aligned}$$

Set the right-hand side of the above inequality equal to 0, we get

$$0 < P\left(1+\underline{\sigma}'_t, e^k_t\right) - P\left(1-\underline{\sigma}'_t, e^k_t\right) = \frac{(1-\pi)(1-\theta)w - \delta(1-\pi)\Delta W^K}{\pi\theta w + \delta\pi\Delta W^K} < 1$$

Since $\left(2EU^K_{(1)} - 2EU^K_{(2)}\right)\Big|_{\sigma_t=0} > 0$, $\left(2EU^K_{(1)} - 2EU^K_{(2)}\right)\Big|_{\sigma_t=\underline{\sigma}'_t} \leq 0$, $\frac{\partial\left(2EU^K_{(1)}-2EU^K_{(2)}\right)}{\partial\sigma_t} < 0$,
Then $\exists\underline{\sigma}_t \in (0, \underline{\sigma}'_t] \subset (0, \phi]$ s.t. $\left(2EU^K_{(1)} - 2EU^K_{(2)}\right)\Big|_{\sigma_t=\underline{\sigma}_t} = 0$ and when $\sigma_t < \underline{\sigma}_t$, $EU^K_{(1)} > EU^K_{(2)}$; when $\sigma_t > \underline{\sigma}_t$, $EU^K_{(1)} < EU^K_{(2)}$.

Next, let us check the sign of $2EU^K_{(2)} - 2EU^K_{(3)}$.
By $\Delta W^A \geq \Delta W^K$, we have

$$\begin{aligned}2EU^K_{(2)} - 2EU^K_{(3)} \geqslant &\pi(1-\theta)w - \theta(1-\pi)w\left[P\left(1+\sigma_t, e^K_t\right) - P\left(1-\sigma_t, e^K_t\right)\right]\\ &- \delta\left\{\pi + (1-\pi)\left[P\left(1+\sigma_t, e^K_t\right) - P\left(1-\sigma_t, e^K_t\right)\right]\right\}\Delta W^A\end{aligned}$$

Substituting into the above equation of $\underline{\sigma}'_t$, we get

$$\begin{aligned}\left(2EU^K_{(2)} - 2EU^K_{(3)}\right)\Big|_{\sigma_t=\underline{\sigma}'_t} \geqslant &\left(2-\frac{1}{\pi}\right)(1-\theta)w\\ &+ \delta\left\{\frac{(1-\pi)^2}{\pi}\Delta W^K - \pi\Delta W^A\right\}\end{aligned}$$

Then by $(1-\theta)w - \delta\Delta W^A > 0$, we have

$$\left(2EU^K_{(2)} - 2EU^K_{(3)}\right)\Big|_{\sigma_t=\underline{\sigma}'_t} > 0$$

Moreover, we have

$$\begin{aligned}2EU^K_{(2)} - 2EU^K_{(3)} \leqslant &\pi(1-\theta)w - \theta(1-\pi)w\left[P\left(1+\sigma_t, e^K_t\right) - P\left(1-\sigma_t, e^K_t\right)\right]\\ &- \delta\left\{\pi + (1-\pi)\left[P\left(1+\sigma_t, e^K_t\right) - P\left(1-\sigma_t, e^K_t\right)\right]\right\}\Delta W^K\end{aligned}$$

Set the right-hand side of the above inequality equal to 0, we get

$$0 < P\left(1+\overline{\sigma}'_t, e^K_t\right) - P\left(1-\overline{\sigma}'_t, e^K_t\right) = \frac{\pi(1-\theta)w - \delta\pi\Delta W^K}{(1-\pi)\theta w + \delta(1-\pi)\Delta W^K} < 1$$

Since $\left(2EU^K_{(2)} - 2EU^K_{(3)}\right)\Big|_{\sigma_t=\underline{\sigma}'_t} > 0$, $\left(2EU^K_{(2)} - 2EU^K_{(3)}\right)\Big|_{\sigma_t=\overline{\sigma}'_t} \leq 0$, $\frac{\partial\left(2EU^K_{(2)}-2EU^K_{(3)}\right)}{\partial\sigma_t} < 0$,
Then $\exists\overline{\sigma}_t \in (\underline{\sigma}'_t, \overline{\sigma}'_t] \subset [\underline{\sigma}_t, \phi]$ s.t. $\left(2EU^K_{(2)} - 2EU^K_{(3)}\right)\Big|_{\sigma_t=\overline{\sigma}_t} = 0$ and when $\sigma_t < \overline{\sigma}_t$, $EU^K_{(2)} > EU^K_{(3)}$; when $\sigma_t > \overline{\sigma}_t$, $EU^K_{(2)} < EU^K_{(3)}$.

Hence, under parameter space 3, given any e^K_t, there exists $(\underline{\sigma}_t, \overline{\sigma}_t)$ such that

- When $\sigma_t < \underline{\sigma}_t$, there is a unique equilibrium where $r^H_t = r^L_t = \eta^H$, and the general never coups.
- When $\underline{\sigma}_t \leq \sigma_t < \overline{\sigma}_t$, there is a unique equilibrium where $r^H_t = \eta^H, r^L_t = \eta^L$, and the general coups if he is a high type but receives a low signal.
- When $\sigma_t > \overline{\sigma}_t$, there is a unique equilibrium where $r^H_t = r^L_t = \eta^L$, and the general coups if he is a high type.

□

Proposition B.6.3 (Power Evolution under Meritocracy)**.** *Consider the parameter space where meritocracy endogenously arises. There exist two threshold values* $(\underline{H}^K, \overline{H}^K)$*, such that*

- *When* $e_t^K > \overline{H}^K$, *the ruler has a probability greater than* $1/2$ *to witness his power to grow over time.*
- *When* $e_t^K < \underline{H}^K$, *the ruler has a probability greater than* $1/2$ *to witness his power to decline over time.*

Proof. Omitted. A natural implication of the above propositions. □

C Additional Information on the Seshat Databank

This section presents additional details and summary statistics for the Seshat Global History Databank used in the empirical analysis in Section 6.

C.1 Background Information

Original Purpose. The Seshat project was created to systematically organize dispersed historical and archaeological knowledge so that scholars can empirically test competing theories about cultural evolution and long-run historical dynamics (notably the evolution of complex societies and inequality). The databank is explicitly positioned as a theory-neutral infrastructure to test rival explanations (Turchin et al., 2015).

Intercoder Reliability/Quality Control. Turchin et al. (2015) details procedures for coding consistency and validation: (i) variable-specific coding sheets and a codebook developed with subject-matter experts; (ii) a wiki-integrated validation/export pipeline that checks entries against an RDF schema and flags non-compliant or ambiguous codes for editors and coders to correct; and (iii) editorial moderation plus redundancy/proxy measures to handle missing or uncertain evidence.

Data Limitations. The Seshat team explicitly acknowledges that the databank is a work in progress and that its initial World Sample-30 (i.e., 30 areas around the world, later on expanded to 35 areas) is a stratified, non-exhaustive sample. They chose a geo-temporal sampling strategy for the first phase and plan to incorporate cultural/ethnic proximity as coverage expands. They note that historical and archaeological records are incomplete, so variables are coded with redundancy and proxies (e.g., largest city size, territorial extent as population proxies). They emphasize transparency and robustness, as definitions and coding assumptions are made explicit. However, despite the best effort, coding and missing data issues may still be present, and researchers are encouraged to test whether conclusions hold under alternative ways of coding (Turchin et al., 2015).

Natural Geographic Area (NGAs). Natural Geographic Area (NGA) is the basic spatial sampling unit in Seshat. An NGA is a fixed map-drawn polygon, typically of the order of 100×100 km, though sizes may vary, that does not change over time. NGAs provide a stable framework for coding and comparing variables over long historical periods. The Seshat Databank samples from 35 geographically and developmentally representative NGAs worldwide (Turchin et al., 2015).

Polity. In Seshat, a polity is an independent political unit ranging from villages and chiefdoms to states and empires that possesses sovereignty (i.e., it is not subordinate to a higher authority). It is linked to the territories (e.g., NGAs) it controls, and because polities evolve over time, Seshat treats them as dynamic entities, often splitting long sequences into successive polities when major transformations occur (e.g., Roman Regal; Early/Middle/Late Republic; Principate; Dominate) (Turchin et al., 2015).

C.2 Coding of Variables

Table 1 below provides the coding definitions of the Seshat variables used in the empirical analysis in Section 6 of the paper.

To illustrate how specific polities are classified as having merit promotion or not, Table 2 provides several examples along with Seshat expert notes explaining the classification decisions.

Table 1: Coding Definitions for Seshat Variables used in Section 6

Variable	**Coding Rule / Definition**
Merit Promotion	Code *present* (=1) if there are regular, institutionalized procedures for promotion based on performance.
Hereditary Elite	Code *present* (=1) if members of the "elite" inherit their status and positions; if the ruler's position is inherited most of the time, that is sufficient to code this variable as present.
Executive Constraints by Government	Code *present* (=1) if government officials (e.g., judiciary/legislature) can veto or overturn executive decisions (including removing a political appointee), or withhold cooperation (e.g., refuse funds or raising troops), regardless of whether such limits were actually exercised.
Executive Constraints by Non-Government	Code *present* (=1) if non-governmental organizations (elite, social, community, or economic groups, etc.) can veto or overturn executive decisions (including removing a political appointee), or withhold cooperation (e.g., refuse funds or raising troops), regardless of whether such limits were actually exercised.
Executive Constraints by Impeachment	Code *present* (=1) if there is an official procedure to impeach the ruler, regardless of whether such limits were actually exercised.
At Least One Executive Constraint	Constructed indicator set to 1 if either "Executive Constraints by Government" or "Executive Constraints by Non-Government" or "Impeachment" is *present*.

Notes: Binary indicators as used in Section 6. Definitions summarized from the Seshat Code Book.
Source: `https://seshat-db.com/variable-hierarchy/`

Table 2: Examples of Meritocratic vs. Non-Meritocratic Polities in Seshat

Polity (Date)	**Merit Promotion**	**Seshat Expert Rationale (abridged)**
Asuka (538–710 CE)	Present	System of caps and ranks favored appointments and promotions based on merit.
Oyo Empire / Ilú-oba O`yo´ (1601–1835 CE)	Present	Seventy *Eso* war-chiefs (senior/junior titles); titles not hereditary, conferred individually on merit for military efficiency.
Kingdom of Hawai'i, Kamehameha period (1778–1819 CE)	Present	Governors owed appointment to executive ability and tested loyalty to the king rather than chiefly rank.
Norman England (1066–1153 CE)	Absent	Administrative and military offices awarded by noble birth/lineage and personal connections; feudal obligations and loyalty.
Isin–Larsa (2004–1763 BCE)	Absent	Father-to-son transmission created closed corporations; elites monopolized access to posts and technical knowledge.
Magadha – Maurya Empire (324–187 BCE)	Absent	A Caste system. Despite efforts to select competence, offices became virtually hereditary; quality of officials declined over time.

Notes: Entries summarized from expert coding notes on the Seshat official website.
Source: `https://seshat-db.com/sc/merit_promotions_all/`

C.3 Summary Statistics and Robustness Checks

The Seshat Global History Databank comprises a panel of 291 historical polities drawn from 35 geographically and developmentally representative *Natural Geographic Area* (NGAs) worldwide.[3]

The following figure plots the location of the NGAs.

[3]For the most recent version of the database, check `https://seshat-db.com`.

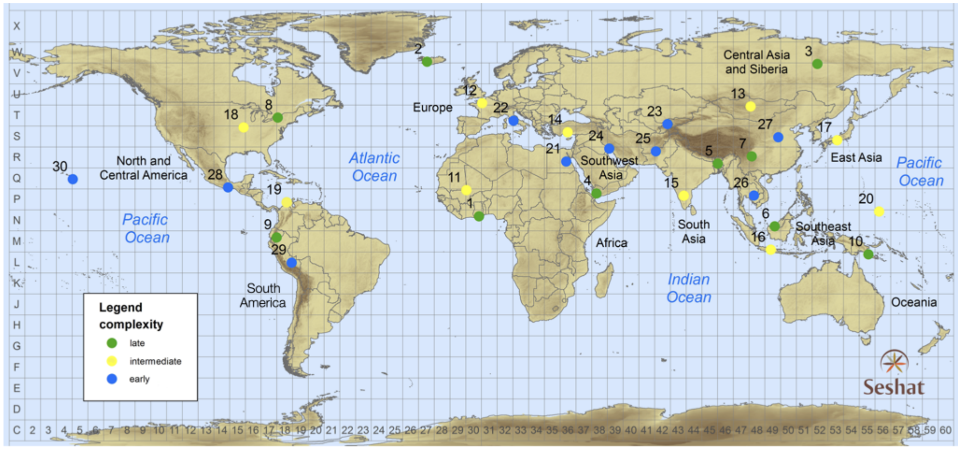


Figure C.1: Location of the Natural Geographic Areas in Seshat Database

The summary statistics are given in the following table.

Table 3: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Number of NGA	35				
Number of Polity	291				
Century	4000 BCE - 1900 CE				
Merit Promotion	750	0.298667	0.457979	0	1
Hereditary Elite	961	0.919355	0.268095	0	1
Exec Const by Govt	363	0.421488	0.494479	0	1
Exec Const by Non-Govt	335	0.489552	0.500639	0	1
Exec Const by Impeachment	423	0.113475	0.317548	0	1
At least one Exec Const	486	0.475309	0.499905	0	1

The following table provides robustness checks for Table 1 in the main text. The first column performs the benchmark regression with robust standard error. Column (2) excludes East Asia, while Column (3) examines the subsample before the eighteenth century. The results in the first three columns are similar to those in the main results. Column (4) uses standard error clustered at the NGA level, and the coefficient for *Merit Promotion* becomes less significant. This is a general feature of the Seshat Database. Because the polities covered in the current version of the database were drawn primarily from only 35 NGAs worldwide, NGA-level clustering tends to absorb a considerable portion of the statistical power. In comparison, spatial correlation is less of a concern, since most polities are not adjacent or even close to each other. Nonetheless, the coefficient for hereditary elites remains positive and statistically significant throughout the permutations.

Table 4: Robustness Checks

VARIABLES	(1) Exec Const	(2) Exec Const	(3) Exec Const	(4) Exec Const
Merit	-0.249**	-0.354***	-0.192*	-0.249
	(0.106)	(0.105)	(0.112)	(0.157)
Hereditary	0.814***	0.935***	0.962***	0.814***
	(0.170)	(0.152)	(0.168)	(0.107)
FE	NGA-Century	NGA-Century	NGA-Century	NGA-Century
Sample	All	Excl. East Asia	<17 century	All
Error	Robust	Robust	Robust	Clu NGA
Observations	299	229	274	299
R-squared	0.505	0.661	0.498	0.505

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. The dependent variable is *At least One Executive Constraint*, which is the same as the dependent variable in the main text Table 1 Column (1) and (2).

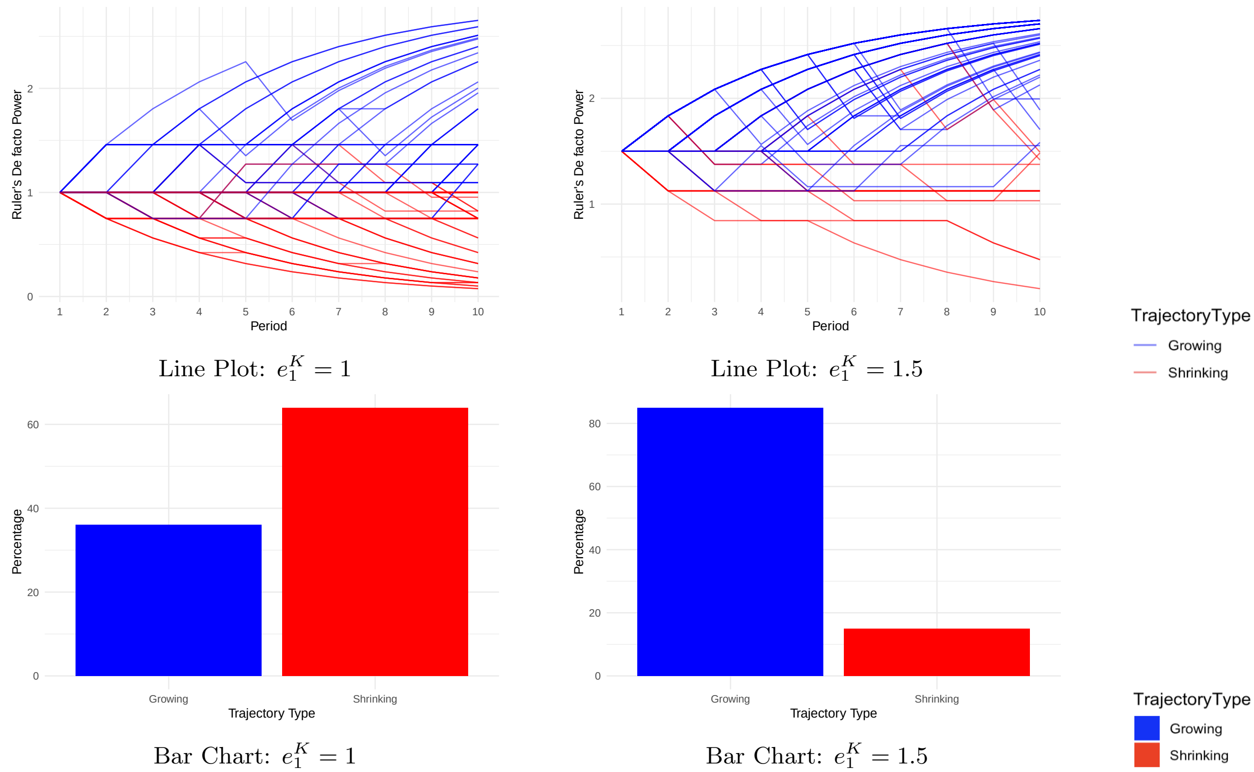
D Simulation of Dynamic Game

D.1 Exogenous Meritocracy

This section provides simulations of the dynamic game under ignorance about reinforcement, with true meritocracy always arises. Assume the conflict function takes the form: $P(x, y) = \frac{x^3}{x^3+y^3}$. Set parameter values as: $\mu = \frac{1}{4}$, $\delta = \frac{3}{4}$, $\theta = 0.9$, $\pi = 0.7$, $w = 4$, $Z = 3$, $e^K_{t=1} = 1$ or 1.5. One hypothetical scenario is consisted of 10 generations (periods) of game play, with 100 scenarios being simulated for each $e^K_{t=1}$. For each scenario, if $e^K_{t=10} > e^K_{t=1}$, then it is categorized as a "Growing Trajectory"; otherwise, it is categorized as a "Shrinking Trajectory".

D.2 Endogenous Meritocracy

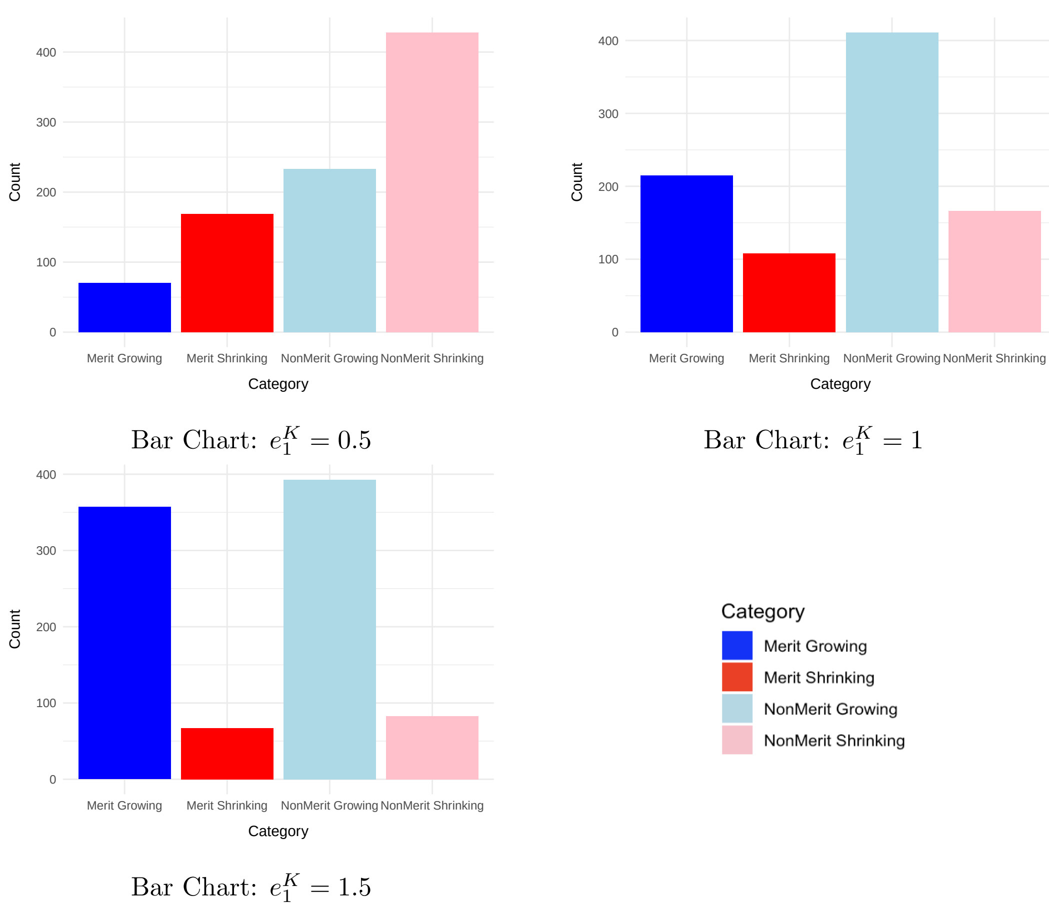
This section provides simulations of the dynamic game under ignorance about reinforcement, with true meritocracy endogenously arising in period t if $\underline{\sigma}_t \leq \sigma_t \leq \overline{\sigma}_t$. Assume the conflict function takes the form: $P(x, y) = \frac{x^3}{x^3+y^3}$. Set parameter values as: $\mu = \frac{1}{4}$, $\delta = \frac{3}{4}$, $\theta = 0.9$, $\pi = 0.7$, $w = 4$, $Z = 3$, $\phi = 0.9$, $e^K_{t=1} = 0.5$ or 1 or 1.5. One hypothetical scenario is consisted of 10 generations (periods) of game play, with 100 scenarios being simulated for each $e^K_{t=1}$. For each period t s.t. $t > 1$, if $e^K_t > e^K_{t-1}$, then it is categorized as a "Growing" period; otherwise, it is categorized as a "Shrinking" period. Moreover, a period t is considered under meritocratic institution (denoted as "Merit") if true meritocracy endogenously arises in this period; otherwise, it is considered under non-meritocratic institution (denoted as "NonMerit").



Line Plot: $e_1^K = 1$ Line Plot: $e_1^K = 1.5$

Bar Chart: $e_1^K = 1$ Bar Chart: $e_1^K = 1.5$

Figure D.1: Trajectories of Ruler's Power under Exogenous Meritocracy



Bar Chart: $e_1^K = 0.5$ Bar Chart: $e_1^K = 1$

Bar Chart: $e_1^K = 1.5$

Figure D.2: Categories of Ruler's Power Growth under Endogenous Meritocracy

E Additional Discussion on Power

What is power? We use the term "power" to broadly capture the *de facto* capacity of a political actor to prevail in confrontation by mobilizing various means including (i) coercive instruments (e.g., military and internal security forces), (ii) organizational reach (e.g., bureaucracies and administrative routines), and (iii) fiscal/extractive resources (e.g., revenues and budgets). At the state level, Mann (1986) distinguishes despotic from infrastructural power, the latter being the logistic capacity to penetrate society and implement decisions through social organizations. Complementing this, Tilly (1992) emphasizes the coercive and extractive foundations of state building under war making. Weber (1978) links durable domination to rational–legal authority exercised via professional bureaucracy. Finally, Levi (1988) stresses that rulers' ability to raise revenue — often secured by quasi-voluntary compliance 'in the shadow of coercion' — is central to their capacity. Therefore, we define power as the *de facto* coercive, organizational, and extractive capacity that enables an actor to prevail in a political conflict.

Power in our model. In our model, we represent power explicitly in the conflict function $P(e, e^K)$: the general's power is e, and the ruler's is e^K. A higher e^K shifts $P(\cdot)$ so that any given general is less likely to prevail in a confrontation. Substantively, as discussed above, power can draw on three forms of capacity: *coercive*, *organizational*, and *extractive*. Each of these forms can in turn originate from two sources. *Institutional/infrastructural power*, derived from organized coercive apparatuses, established bureaucracies, and centralized revenues, tends to persist and can be transmitted across generations. By contrast, *personal power*, derived from an individual's military talent, charisma, and personal networks, is typically transitory and less inheritable. In the static setting, where institutional conditions can be treated as fixed in the short run, distinguishing between these two sources of power makes little difference. However, in the long run, the distinction becomes crucial, which is why we formally separate institutional and personal components only in the dynamic section of the model.

How does power transfer to a second generation? Leadership succession is a canonical stress test of *de facto* power. When organizational assets are institutionalized rather than personalized, they can be passed on to successors, preserving e^K across generations. In our framework, this corresponds to the partial durability of institutional power: when coercive, organizational and fiscal capacities are embedded in bureaucracies and coercive apparatuses, e^K remains high for successors. By contrast, when political power is highly personalized, its succession is more uncertain, effectively increasing σ_t in our model. Observationally, successful intergenerational transfer of e^K should be associated with (i) continuity in centralized control over appointments and budgets, (ii) effective monitoring that maintains the informativeness of signals, and (iii) continued reliance on signal-contingent rewards.

References

Levi, M. (1988). *Of Rule and Revenue.* Berkeley: University of California Press.

Mann, M. (1986). *The Sources of Social Power, Vol. I: A History of Power from the Beginning to A.D. 1760.* Cambridge: Cambridge University Press.

Tilly, C. (1992). *Coercion, Capital, and European States, AD 990-1992.* Cambridge, MA: Blackwell.

Turchin, P., R. Brennan, T. E. Currie, K. Feeney, P. François, H. Whitehouse, et al. (2015). Seshat: The global history databank. *Cliodynamics 6*(1), 77–107.

Weber, M. (1978). *Economy and Society.* Berkeley: University of California Press.