On the Non-Emergence of Limited Government: Private Information and Institutional Choice*

Mark Dincecco † David Rahman ‡

19 December 2005

Abstract

Why is limited government so scattered over time and place? We examine how private information that concerns national security affairs affects the choice of political institutions. To do so, we model limited government as a costly institutional alternative through which a ruler may transmit a credible signal of impending foreign attack to its subject. The ruler must decide whether to erect limited government or not. We show that limited government – though optimal in terms of total welfare – does not necessarily emerge to solve private information dilemmas. Indeed, there exist a diverse set of outcomes depending on the probability of foreign attack and the ruler's building cost. In the one-shot version, the ruler may forego limited government and exploit the subject some or all of the time. Repeated interactions may induce honest behavior by the ruler without limited government.

JEL Classification: C70, P16, N20

---

*Mark Dincecco would like to thank Jean-Laurent Rosenthal for his guidance, dedication, and generosity. He is also grateful for helpful suggestions made by Michael Chwe, Christian Hellwig, and David Levine. Lastly, he would like to thank UCLA and the IHS for financial support.

†Corresponding author: University of California, Los Angeles (UCLA), Department of Economics, Box 951477, Los Angeles, CA 90095; email: mdincecc@ucla.edu

‡University of California, San Diego (UCSD), Department of Mathematics, 9500 Gilman Drive, La Jolla, CA 92093; email: drahman@euclid.ucsd.edu

1 Introduction

The great confusion over Iraqi weapons of mass destruction (WMD) in 2003 highlights the role of private information in national security affairs. President Bush alleged that Saddam Hussein possessed a large cache of WMD that he wished to use to attack the United States, but no weapons were ever found. Was this an intelligence failure, or did someone lie? Hussein may have fooled the CIA director – and in turn the President – to believe that Iraqi WMD were a "slam-dunk." Or, acting on his own accord, the President may have exaggerated the WMD threat to the public. Regardless, the WMD debate illustrates an important asymmetry: high-ranking government officials such as the CIA Director and the President possess information about national security affairs that the public does not possess. What matters is if and how government officials choose to share information with the public. Officials may use this informational edge to society's advantage or not.

In this paper, we examine how private information about national security affairs affects the choice of political institutions. Our work draws on Green (1993), who models a ruler and subject that debate over the allocation of the crop that the subject grows. Private information arises through a third "agent," the dragon, who represents the threat of foreign (e.g., terrorist) attack.[1] Only the ruler can see from the turret of its castle if the dragon – capable of eating the entire crop – is coming or not. The ruler can kill the dragon by wielding a lance, but to do so it must eat amount z of the subject's crop. Yet, for the subject to yield z, the ruler must first provide a credible signal that the dragon is indeed coming. Otherwise, the subject will yield just $z-1$ of its crop to the ruler. The ruler provides a credible signal only by dragging its lance to the subject's field, a strenuous activity that the ruler would not take unless the threat of dragon attack was real. The subject observes this credible signal and, in turn, allows the ruler to eat z of its crop for strength to slay the dragon.

This exercise illustrates the roles that predation and veto right play in the establishment of limited government, which we define as:

"...a form of government which includes a systematic arrangement to obtain the

[1] The issue of private information in matters of government is long-standing. Green argues that in medieval times English monarchs had superior access to diplomatic resources (i.e., informants) within the French court able to provide advance notice of attack. Hence, with the loss of Normandy as a buffer between England and France in the 13th century, English barons had to rely on the king to inform them of threats of French invasion.

> consent of the governed to the actions of the government...that requires state action to be based on mutually voluntary, two-way communication between the ruler and the subjects." [2]

If the subject was unable to veto the ruler's demand for amount z of the crop, then the ruler would predate by claiming in all circumstances that the dragon was coming (and thereby consume z rather than $z-1$ of the crop). The fact that the subject is able to veto ruler demands prevents this sort of exploitation, for the ruler prefers to "tie its own hands." By doing so, the ruler protects itself from obliteration when the dragon actually attacks.

Indeed, the limited government outcome maximizes total welfare. If the ruler does not drag the lance to the subject's field, then the subject will not let him eat amount z of the crop. The ruler will not have enough stamina to fend off the dragon, who will in turn ravish both itself and the subject. Thus, the subject prefers to yield z of the crop to the ruler when the dragon comes.

This result suggests that limited government is the optimal institutional solution where private information about national security affairs exists.[3] Private information environments, however, are common. Why, then, is limited government so scattered over time and place? To answer, we model in an explicit way institutional choice by the ruler, who must decide whether to erect limited government or not in order to combat the threat of foreign attack. This technical innovation highlights the difference between the optimal outcome and the ruler's preferred outcome. What drives institutional choice is not that which maximizes total welfare but that which yields the greatest possible utility to the ruler. Often times the ruler does not wish to "build" limited government, regardless of its benefit to society.

Incorporation of institutional choice generates a diverse set of outcomes. We find that limited government is not a panacea. In fact, limited government is not "worth" it to the ruler if foreign invasion is unlikely. Rather, it will forego it and bear the (i.e., small) risk that foreign attack brings. The ruler will also pass up limited government when there is a high probability of foreign invasion. In this case, it will exploit the subject's fear of attack to predate and demand high (versus low) taxes all of the time. If the ruler finds limited government too costly, then it will never erect it, regardless of the chance of foreign invasion.

[2] We quote Green, page 2.

[3] McGuire and Olson (1996), on the other hand, present a model where limited government never arises. Rather, an autocrat rules indefinitely. In this environment, however, there is no private information, which explains why institutional change does not occur. To gain additional resources, the autocrat does not need to credibly convey threat alerts to his subjects, who are able to recognize them themselves. This highlights the crucial role that private information plays in limited government outcomes.

Repeated interactions may induce honest behavior by the ruler to elicit high taxes from the subject without limited government.

This paper is related to the literature on the origins of democracy.[4] Acemoglu and Robinson (2000) argue that franchise extension enables autocratic elites to preempt destructive revolution by committing them to future redistribution. Lizzeri and Persico (2003), however, claim that fear of revolution by elites is not sufficient cause for democratization. Elites choose to extend the franchise to encourage politicians to pursue policies that promote the greater public good – and benefit elites in turn – rather than to seek patronage (i.e., pork-barrel politics). Both sets of authors employ their models to explain the rise of limited government within 19th century European states. This work, in contrast, investigates the many instances past and present where limited government fails to emerge. To do so, we incorporate private information about national security affairs, which enables the government in many instances to exploit its subjects indefinitely. Our results suggest that we must add this element, in addition to fear of revolution and distaste for pork-barrel politics, to the list of factors that promote limited government reform (or lack thereof).

The plan of the paper is as follows. In section 2, we set up the model. In section 3, we present the results of the one-shot version. In section 4, we analyze an infinitely repeated version. We investigate real-world implications of the model in section 5. An appendix with proofs follows.

2 Model Set-up

2.1 Agents and Information

There are two agents, ruler R and subject S.

Private information concerns the state of the world X, where $X = 0$ or $X = 1$. Nature N determines this state. With $X =$ foreign invasion, $X = 0$ means that there will be no foreign invasion (i.e., peacetime) and $X = 1$ means that foreign invasion will occur.

Let $p = \Pr(X = 1)$ and $(1 - p) = \Pr(X = 0)$. We may think that foreign invasion (i.e., $X = 1$) is uncommon and that peacetime (i.e., $X = 0$) represents the status quo. We solve the game and draw conclusions over the entire range of p values $0 \leq p \leq 1$, however.

[4] There are many works that explore the varied roles that political institutions play in economic outcomes. See, among others, North (1981), Olson (1982), North and Weingast (1989), Acemoglu, Johnson, and Robinson (2001), Stasavage (2003), Aghion, Alesina, and Trebbi (2004), and Acemoglu (2005).

Private information arises as follows: the ruler knows whether or not foreign attack will occur at the start of the game, but the subject does not know until the end of the game. Thus, the subject must rely on the ruler to indicate the true state of the world. The utility rankings over outcomes are such, though, that the ruler has an incentive to lie about the true state so as to expropriate more from the subject than otherwise possible.

The game revolves around the ruler's request for high (H) or low (L) taxes and the subject's response to this request. The subject may acquiesce (A) or fight (F). When foreign attack will occur, the ruler must secure high taxes from the subject to mount a military defense to thwart it. Otherwise, both the ruler and the subject will suffer dreadfully. There are three ways that the ruler may attain high taxes:

- **Credible Signal**: The ruler shows proof, or provides a credible signal (CS), of foreign invasion. To do so, the ruler must first build (B) a credible-signal institution, which represents limited government: "... a form of government ... that requires state action to be based on mutually voluntary, two-way communication between ruler and subject." To build and make use of the credible-signal institution is a sure-fire way for the ruler to elicit high taxes from the subject. The total cost to erect the credible-signal institution is $I = I_R + I_S$, where I_R represents the portion of the building cost borne by the ruler and I_S the portion borne by the subject, with $I_R, I_S \geq 0$.

- **State-Contingent Request**: When foreign invasion is imminent, the ruler may simply request high taxes (H) rather than build and make use of the credible-signal institution. During peacetime, the ruler will request low taxes (L). Though the ruler is honest about the true states of the world, the lack of a credible signal means that the subject is unsure whether the ruler is telling the truth about foreign attack or that there is peace and the ruler wishes to lie and exploit him. Hence, the subject may fight high tax requests by the ruler. This stands in contrast to the previous option, for if the ruler builds and makes use of the credible-signal institution, then the subject will certainly acquiesce to high taxation.

- **Non-State-Contingent Request**: The ruler may indeed attempt to predate by lying to the subject about the threat of foreign attack. Predation amounts to a request by the ruler for high taxes regardless of the true state of the world, including during peacetime.

2.2 Timing

As figure 1 illustrates, the timing of the game is as follows:

1. The ruler chooses whether to build (B) or not build (NB) the credible-signal institution. If the ruler plays B, then the upper subgame is played, and if the ruler plays NB, then the lower subgame is played. As mentioned, the total cost of the credible-signal institution is $I = I_R + I_S$, where I_R represents the portion of the building cost borne by the ruler and I_S the portion borne by the subject, with $I_R, I_S \geq 0$. Cost I is incurred as soon as the ruler decides to build the credible-signal institution (i.e., once the ruler plays B). Once built, the ruler may not revert out of the upper subgame.

2. The state of the world is revealed to the ruler (and not to the subject): $X = 0$ (i.e., peacetime) or $X = 1$ (i.e., impending foreign attack).

3. Given this information, and depending on whether or not the ruler has chosen to erect the credible-signal institution (i.e., whether the upper or lower subgame is played), the ruler chooses to provide a credible signal of foreign invasion CS, request high taxes H without offering a credible signal, or request low taxes L.

4. The subject reacts to the ruler's announcement by acquiescing A or fighting F. The subject is always successful if it chooses to fight.

5. The state of the world is revealed to the subject. The ruler and the subject receive payoffs for their actions.

2.3 Strategies

The ruler's strategy takes the form $s_R = (j, k_B, l_B, k_{NB}, l_{NB})$. The j component refers to whether or not the ruler builds the credible-signal institution, B or NB. The k_B and l_B components refer to the ruler's respective actions at $X = 1$ and $X = 0$ given that the ruler has built the credible-signal institution. At both k_B and l_B the ruler may offer a credible signal CS that foreign invasion will occur or simply request high taxes H or low taxes L. The k_{NB} and l_{NB} components refer to the ruler's respective actions at $X = 1$ and $X = 0$ given that the ruler has chosen not to build the credible-signal institution. The ruler may request high taxes H or low taxes L at both k_{NB} and l_{NB}. Of course, the ruler cannot offer a credible signal CS at k_{NB} or l_{NB} because it has chosen not to build the credible-signal institution.

For example, the strategy $(B, CS, L, \cdot, \cdot)$ means that the ruler builds the credible-signal institution B (i.e., the j component), provides a credible signal CS that $X = 1$ when in fact $X = 1$ (i.e., the k_B component), and asks for low taxes L at $X = 0$ (i.e., the l_B component). By building the credible-signal institution the ruler and the subject are in the upper subgame. Therefore, the k_{NB} and l_{NB} components do not pertain. As a convention, we use $(\cdot, \cdot)$ to symbolize this portion of the ruler's strategy.

The subject's strategy takes the form $s_S = (\alpha_l, \alpha_u, \gamma)$, where α_l, α_u, and γ are the three information sets at which the subject plays.[5] We use the strategy $(\cdot, F, A)$ as an example. To begin, notice that only the latter two components of the subject's strategy, α_u and γ, pertain. This indicates that the credible-signal institution has been built, and the ruler and the subject are in the upper subgame. Hence, at information set α_u, the subject fights F high tax requests because the ruler does not provide a credible signal CS that indeed $X = 1$. When the ruler provides a credible signal CS, then we are at information set γ and the subject acquiesces A. As we have mentioned, the lower subgame component α_l does not apply. Therefore, $(\cdot)$ symbolizes this portion of the subject's strategy.

2.4 Utility Rankings

As figure 1 indicates, the total game is comprised of lower and upper subgames. We first describe outcomes and utility rankings over them for the lower subgame. We then do the same for the upper subgame. Outcomes and utility rankings are very similar for both. After, we describe outcomes and corresponding utility rankings for the game in total by comparing lower and upper subgames.

2.4.1 Lower Subgame

During peacetime (i.e., $X = 0$), outcomes for the lower subgame are: c_{NB} if the subject fights the ruler's request for high taxes; d_{NB} if the subject acquiesces to the ruler's request; and e_{NB} if the ruler asks for low taxes.[6] Similarly, when foreign invasion will occur (i.e., $X = 1$), outcomes are: a_{NB} if the subject fights the ruler's request for high taxes; b_{NB} if the subject acquiesces to the ruler's request; and a_{NB} if the ruler asks for low taxes.

The ruler's utility rankings over outcomes in the lower subgame, where it has opted not

[5] As noted, we assume that the subject is always successful if it chooses to fight F.

[6] The subject will always acquiesce to the ruler's request for low taxes. This is an implicit simplification since the subject does not have an action when the ruler requests low taxes L.

to build the credible-signal institution, are:

$$U_R(d_{NB}) > U_R(b_{NB}) > U_R(e_{NB}) > U_R(c_{NB}) > U_R(a_{NB})$$

The ruler derives the greatest utility, $U_R(d_{NB})$, from predation on the subject (i.e., where the subject acquiesces). On the other hand, the ruler derives the lowest utility, $U_R(a_{NB})$, if the subject fights its request for high taxes when foreign attack will occur, since foreign invaders will ravish both ruler and subject.

The subject's utility rankings in the lower subgame are:

$$U_S(e_{NB}) > U_S(b_{NB}) > U_S(c_{NB}) > U_S(d_{NB}) > U_S(a_{NB})$$

These payoffs indicate that the subject does best, $U_S(e_{NB})$ when in peacetime it pays low taxes. The subject suffers, on the other hand, when it allows the ruler to predate on him, $U_S(d_{NB})$, but not so badly as when it fights high tax requests by the ruler when foreign attack will occur, $U_S(a_{NB})$.

2.4.2 Upper Subgame

As figure 1 also shows, outcomes are similar for the upper subgame as for the lower one: c_B if the subject fights the ruler's request for high taxes during peacetime (i.e., $X = 0$); d_B if the subject acquiesces to the ruler's request; and e_B if the ruler asks for low taxes; a_B if the subject fights the ruler's request for high taxes if foreign invasion will occur (i.e., $X = 1$); b_B if the subject acquiesces to the ruler's request; and a_B if the ruler asks for low taxes. There is also a new set of outcomes that occur when the ruler plays the credible-signal strategy. Of these, we wish to highlight f_B, where the ruler provides a credible signal that foreign attack will occur and, in turn, the subject acquiesces to high taxation.

Recall, though, that we must deduct building costs I_R and I_S, respectively, from all upper subgame payoffs, where the total cost to erect the credible-signal institution is $I = I_R + I_S$, with $I_R, I_S \geq 0$.[7] For simplicity, we assume that $U_i(a_B) = (U_i(a_{NB}) - I_i)$, $U_i(b_B) = (U_i(b_{NB}) - I_i)$, $U_i(c_B) = (U_i(c_{NB}) - I_i)$, $U_i(d_B) = (U_i(d_{NB}) - I_i)$, and $U_i(e_B) = (U_i(e_{NB}) - I_i)$, for $i = R$ or S.

Hence, the ruler's utility rankings over outcomes in the upper subgame, where it has

[7] We assume that I_R and I_S are never large enough to change the utility rankings over outcomes that I specify.

elected to build the credible-signal institution, are:

$$U_R(d_B) > U_R(b_B) > U_R(e_B) = U_R(f_B) > U_R(c_B) > U_R(a_B)$$

Which, based on the aforementioned assumption, may be rewritten as:

$$(U_R(d_{NB}) - I_R) > (U_R(b_{NB}) - I_R) > (U_R(e_{NB}) - I_R) = U_R(f_B) > (U_R(c_{NB}) - I_R) > (U_R(a_{NB}) - I_R)$$

As for the lower subgame, the ruler does best through predation on the subject, $U_R(d_B)$, where the subject acquiesces. Whenever foreign invasion will occur, however, the ruler prefers to play the credible-signal strategy and receive high taxes for sure rather than simply ask for (and receive) low taxes; $U_R(f_B) > U_R(a_B)$. Again, the ruler suffers most, $U_R(a_B)$, if the subject fights its requests for high taxes when foreign attack will occur.

For the subject:

$$U_S(e_B) > U_S(b_B) = U_S(f_B) > U_S(c_B) > U_S(d_B) > U_S(a_B)$$

Equivalently:

$$(U_S(e_{NB}) - I_S) > (U_S(b_{NB}) - I_S) = U_S(f_B) > (U_S(c_{NB}) - I_S) > (U_S(d_{NB}) - I_S) > (U_S(a_{NB}) - I_S)$$

Whenever foreign invasion will occur, the subject also prefers that the ruler play the credible-signal strategy CS (to which the subject then acquiesces and pays high taxes) to any play for low taxes; $U_S(f_B) > U_S(a_B)$. Again, the subject suffers from predation by the ruler, $U_S(d_B)$, but not so badly as when it fights high tax requests when foreign attack will occur, $U_S(a_B)$.

2.4.3 Game in Total

The ruler's utility rankings over outcomes for the game in total are:

$$U_R(d_{NB}) > U_R(d_B) > U_R(b_{NB}) > U_R(b_B) > U_R(e_{NB}) > U_R(e_B) = U_R(f_B) >$$
$$U_R(c_{NB}) > U_R(c_B) > U_R(a_{NB}) > U_R(a_B)$$

Similarly, total game utility rankings over outcomes for the subject are:

$$U_S(e_{NB}) > U_S(e_B) > U_S(b_{NB}) > U_S(b_B) = U_S(f_B) > U_S(c_{NB}) > U_S(c_B) >$$
$$U_S(d_{NB}) > U_S(d_B) > U_S(a_{NB}) > U_S(a_B)$$

Utility rankings described for the lower and upper subgames continue to hold for the game in total.[8]

3 Equilibrium Results

The various equilibria that emerge in the one-shot version depend on the probability of foreign attack $0 \leq p \leq 1$ and the ruler's cost to erect the credible-signal institution $I_R \geq 0$. So long as the ruler's cost I_R is less than or equal to threshold cost level $\overline{I}_R$, then there exist three possible equilibria, each unique over a particular range of p values. To be specific, the mixed (i.e., sometimes predatory) equilibrium holds for $0 \leq p < \widehat{p}$ values, the credible-signal equilibrium holds for $\widehat{p} \leq p < \overline{p}$ values, and the predatory equilibrium holds for $\overline{p} \leq p \leq 1$ values, where $\widehat{p}$ and $\overline{p}$ represent threshold levels of p such that $0 \leq \widehat{p} < \overline{p} \leq 1$. Please see figure 5 for an illustration of these equilibria. If it is costless for the ruler to construct the credible-signal institution, $I_R = 0$, then the mixed (i.e., sometimes predatory) equilibrium disappears and the credible-signal equilibrium holds for all $0 \leq p < \overline{p}$ values. The predatory equilibrium remains for $\overline{p} \leq p \leq 1$ values. If the ruler's cost I_R exceeds the threshold cost level $\overline{I}_R$, however, then the opposite happens: the mixed (i.e., sometimes predatory) equilibrium replaces the credible-signal equilibrium for all $0 \leq p < \overline{p}$ values. The predatory equilibrium continues to hold for $\overline{p} \leq p \leq 1$ values.[9]

We now characterize these three equilibria – credible-signal, predatory, and mixed (i.e., sometimes predatory) – at length. Please see appendix for all proofs.

[8] The ruler most prefers predation (where the subject acquiesces) and receives a higher payoff, $U_R(d_{NB})$, from predation without building the credible-signal institution than from predation after building the credible-signal institution, $U_R(d_B)$, so long as there is a positive cost I_R to build the credible-signal institution. The subject does best when in peacetime he pays low taxes, $U_S(e_{NB})$ and $U_S(e_B)$. Predation harms the subject, $U_S(d_{NB})$ and $U_S(d_B)$, but it does worst when it fights high tax requests by the ruler when foreign attack will occur, $U_S(a_{NB})$ and $U_S(a_B)$.

[9] Please see appendix for $\overline{I}_R$, $\widehat{p}$, and $\overline{p}$ derivations.

3.1 Credible-Signal Equilibrium

The ruler is always tempted to lie and predate on the subject. Hence in the one-shot version there is no separating (i.e., "honesty") equilibrium $\{(NB, \cdot, \cdot, H, L), (A, \cdot, \cdot)\}$ where the ruler's tax policies enable the subject to reliably infer the true state of the world. The subject always fights high tax requests by the ruler when an intermediate chance $\widehat{p} \leq p < \overline{p}$ exists of foreign invasion. In this case, the ruler prefers to build the credible-signal institution and play the credible-signal strategy CS when foreign attack will occur so long as it is not too costly to do so (i.e., $I_R \leq \overline{I}_R$). The ruler asks for low taxes L otherwise. This is the credible-signal equilibrium:

Proposition 1 *When the ruler bears a cost burden $I_R \leq \overline{I}_R$ of building the credible-signal institution, $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ is the unique perfect Bayesian equilibrium for $\widehat{p} \leq p < \overline{p}$, where $I_R, \overline{I}_R > 0$.*

Proposition 1 encompasses Green's model. Over the range of $\widehat{p} \leq p < \overline{p}$ and $I_R \leq \overline{I}_R$ values, the ruler prefers to tie its own hands and restrict predation through limited government. Figure 2 shows a numerical example of the credible-signal equilibrium where $I = 10$ with $I_R = 2$ and $I_S = 8$.

The ruler plays the credible-signal strategy $(B, CS, L, \cdot, \cdot)$ more and more often as it bears less and less of the cost of building the credible-signal institution. That is, the credible-signal equilibrium $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ becomes unique for a wider range of $\widehat{p} \leq p < \overline{p}$ values as the cost burden I_R decreases.

$I_R = 0$ means that the ruler bears none of the cost of erecting limited government (i.e., the subject bears the entire cost $I = I_S$). In turn, we have:

Remark 1 $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ *is the unique perfect Bayesian equilibrium when $p < \overline{p}$ for $I_R = 0$.*

If the credible-signal institution is costless for the ruler to build, then it will erect it even when the probability of foreign invasion is quite low. Figure 3 shows a numerical example of the credible-signal equilibrium where the subject bears the entire cost of building the credible-signal institution; $I_S = I = 10$.

3.2 Predatory Equilibrium

As foreign attack becomes more likely, however, then it is optimal for the subject to always acquiesce to high tax requests by the ruler. The ruler takes advantage of the subject's fear of invasion to exploit him and request high taxes H regardless of whether foreign attack will occur. This results in the predatory equilibrium:

Proposition 2 $\{(NB, \cdot, \cdot, H, H), (A, \cdot, \cdot)\}$ *is the unique perfect Bayesian equilibrium when* $p \geq \overline{p}$ *for all* $I_R \geq 0$.

Hence, we are never certain that the ruler will erect limited government. For high p values, the ruler will predate on the subject and receive high taxes without constructing the credible-signal institution (and hence without providing a credible signal of foreign invasion), regardless of how little it costs the ruler to build.

Figure 4 show a numerical example of the predatory equilibrium where $I_R = 0$.

3.3 Mixed (Sometimes Predatory) Equilibrium

In general, it is not worth it to the ruler to erect limited government for low p values. If foreign invasion is very rare, then the risk of devastation it poses does not merit the cost of constructing the credible-signal institution. Instead, the ruler will mix between predatory and honest behavior and the subject will mix between acquiescing and fighting.

Proposition 3 *When the ruler bears a cost burden* $I_R \leq \overline{I}_R$ *of building the credible-signal institution there exists the unique perfect Bayesian equilibrium in mixed strategies where the ruler does not build (NB) the credible-signal institution, the* R_{1l} *plays* H*,* R_{0l} *mixes over* H *and* L *with* $\tau_0 = \overline{\tau_0}$*, and the subject mixes over* A *and* F *with* $\sigma = \overline{\sigma}$ *for* $p < \widehat{p} < \overline{p}$*, where* I_R*,* $\overline{I}_R > 0$.

We know, however, that for low p values the ruler plays the credible-signal strategy $(B, CS, L, \cdot, \cdot)$ more and more often as it bears less and less of the cost of building the credible-signal institution. Remark 1 indicates that at the extreme (i.e., $I_R = 0$) the ruler does not play the mixed equilibrium strategy at all. On the other hand, the ruler plays the credible-signal strategy $(B, CS, L, \cdot, \cdot)$ less as its cost to build the credible-signal institution grows. This implies the following:

Proposition 4 *So long as the ruler bears more than* $I_R > \overline{I}_R$ *of the cost of building the credible-signal institution, there exists the unique perfect Bayesian equilibrium in mixed strategies where the ruler does not build (NB) the credible-signal institution,* R_{1l} *plays* H*,* R_{0l} *mixes over* H *and* L *with* $\tau_0 = \overline{\tau_0}$*, and the subject mixes over* A *and* F *with* $\sigma = \overline{\sigma}$ *for* $p < \overline{p}$*, where* I_R*,* $\overline{I}_R > 0$.

If erecting the credible-signal institution becomes too costly for the ruler (i.e., $I_R > \overline{I}_R$), then there is no longer any intermediate range of $\widehat{p} < p < \overline{p}$ values over which the ruler constructs it, for the lower bound $\widehat{p}$ becomes zero. Instead, the ruler prefers to play a mixed strategy and risk devastation by foreign attackers for all $p < \overline{p}$ values. For $p \geq \overline{p}$, the predatory equilibrium still holds. Over this range of p values, the ruler's construction cost I_R is irrelevant, since the ruler always prefers to predate on the subject when the likelihood of foreign invasion is high.

4 Repeated Version: Honesty Equilibrium

Does the ruler have to build and make use of the credible-signal institution to convince the subject that foreign invasion is imminent? Recall that, in the one-shot version, there is no separating, or honesty, equilibrium where the ruler's tax policy behavior enables the subject to reliably infer the true state of the world. In the infinitely repeated version, however, the honesty equilibrium may emerge:

Proposition 5 *There exists a discount factor $\delta(p)$ such that for $\delta(p) \geq \overline{\delta}(p)$ the strategy profile for the ruler and the subject with equilibrium behavior $\{(NB, \cdot, \cdot, H, L), (A, \cdot, \cdot)\}$ is a subgame perfect equilibrium of the infinitely repeated game, where $0 \leq \overline{\delta}(p) \leq 1$ and $I_R \geq 0$.*

Proposition 5 indicates that in an infinitely repeated setting the ruler does not need to build and make use of the credible-signal institution to prove to the subject that foreign invasion will occur. If the ruler is patient enough (i.e., $\delta(p) \geq \overline{\delta}(p)$), then repeated interactions themselves may induce it to propose honest tax policies. Figure 6 indicates that when $I_R \leq \overline{I}_R$ the credible-signal equilibrium is just one of four possible outcomes in the infinitely repeated version.[10] Thus, honesty makes for an effective policy alternative to limited government to thwart foreign invasion.[11]

[10] When $I_R = 0$, the credible-signal equilibrium replaces the mixed strategy equilibrium, which fails. When $I_R > \overline{I}_R$, however, the mixed strategy equilibrium replaces the credible-signal equilibrium.

[11] This calls to mind a similar result of McGuire and Olson (1996), where in infinitely repeated setting a "patient" autocrat makes clear to his subjects that he will limit his theft and provide public goods. In turn, his subjects have an incentive to produce, which means that the autocrat obtains more resources for himself than otherwise.

5 Discussion

From the model we learn that limited government does not necessarily emerge in private information environments. Rather, as figure 5 illustrates, it is only built when there is an intermediate likelihood of foreign invasion (i.e., $\widehat{p} < p < \overline{p}$) and when it is not too costly for the ruler to erect (i.e., $I_R \leq \overline{I}_R$).

Limited government is only valuable in the case of foreign attack. However, the ruler must also bear its cost in peacetime, when it is useless. The credible-signal equilibrium fails for low p values (i.e., $p < \widehat{p} < \overline{p}$) for this reason. The weak threat of foreign invasion does not merit the ruler's investment in the credible-signal institution. Instead, the ruler opts to play a mixed strategy and bear the (i.e., small) risk of destruction by foreign invaders. The ruler also chooses not to build the credible-signal institution for high p values (i.e., $p \geq \overline{p}$), for then it is able to exploit the subject's "fear of invasion" and predate for high taxes in all circumstances. In addition, the ruler prefers to forego limited government regardless of the probability of foreign attack (i.e., for all $0 \leq p \leq 1$ values) if it is too costly to build (i.e., $I_R > \overline{I}_R$). Repeated interactions may induce a patient ruler (i.e., $\delta(p) \geq \overline{\delta}(p)$) to propose honest tax policies rather than limited government to defend against foreign invasion.

These results highlight an important distinction between optimal outcomes, which maximize total welfare, and ruler-preferred outcomes, which maximize the ruler's own welfare. Optimal outcomes and ruler-preferred outcomes may overlap. If the cost of the credible-signal institution to the ruler is high and the cost of predation to the subject is low, then the ruler-preferred predatory outcome may generate greater total utility than any other outcome, including the credible-signal one. In this scenario, the predatory equilibrium is both ruler-preferred and optimal.

The ruler-preferred outcome, however, always takes precedence. That is, the ruler chooses the strategy that yields the greatest possible utility to him regardless of its effect on total welfare. To say that limited government is always the outcome that maximizes the ruler's utility is to presuppose a false conclusion.

This observation clarifies certain institutional outcomes in the real world. Limited government emerges where informational asymmetries impose great costs on the ruling government if it does nothing. This does not occur at either extreme. In dictatorial societies, such as North Korea under Kim Jong Il, the public greatly fears an attack by the United States and its allies. Kim Jong Il has no incentive to discourage this fear. Rather, he exploits it to justify high taxation, which he (i.e., purportedly) uses towards military defense. Fear of foreign attack is so great that to secure funds the dictator does not need to provide a credible signal of it through limited government.

We also find institutional inertia at the other end of the spectrum. In Saudi Arabia, the royal family enjoys American military protection and the threat of foreign attack appears low. In turn, the royal family has no reason to erect limited government, for to do so it would require it to yield a great deal of power to others. Limited government is simply not "worth" it to Saudi royals, who reap impressive revenues from petroleum exports under the present institutional regime (i.e., monarchy).

We may speculate further to say that limited government will not readily emerge in any country with secure borders and great wealth in natural resources. The recent history of the Democratic Republic of the Congo (formerly Zaire) illustrates. After independence in 1960, dictator Mobutu Sese Seko seized power. He then ruled with an iron fist for over thirty years. During this period, Mobutu amassed a fortune by looting his country of its export earnings from diamonds. Western powers supported Mobutu's reign as a regional foil against communism.[12] Why in the world would Mobutu favor limited government? A prosperous ruler that does not live in fear of foreign takeover has little incentive to make costly institutional changes that will undermine his authority and income.

At intermediate levels, the threat of foreign attack is not patently obvious, nor is it highly unlikely. Thus, it makes sense that under such circumstances a government must justify its military expenses in a transparent way. Without a credible signal of possible threats, the public will favor lower taxation for defense. As fear of attack mounts, however, the public tends to turn a blind eye to defense spending hikes. To explain, we must extend the results of the model. During the Cold War, Americans justified costly military projects to fend off the Soviet Union. Similarly, after the terrorist attacks of September 11, 2001, Americans were less likely to object to large increases in defense spending (e.g., homeland security), though – excluding Afghanistan – the United States did not enter a military conflict until Operation Iraqi Freedom began in 2003. Ceteris paribus, this amounts to public "acquiescence" to high taxation in the absence of war.

[12] Facts on File World News Digest, 22 May 1997; www.facts.com/wnd/mobutu2.htm

A Proofs

To solve, we break the (i.e., one-shot) game up into its two component pieces, the upper and lower subgames, and examine all equilibrium possibilities. We solve for the total game's equilibria by comparing the results of the two subgames. After, we do the same for the infinitely repeated game.

Please refer to section 2 of the text (i.e., model set-up) and figure 1 for explanations of the notation that we employ.

A.1 Lower subgame

In the lower subgame, pooling, separating, and semi-separating (i.e., mixed) equilibria may arise. We will now exhaust all equilibrium possibilities, eliminating most in the process.

To begin, we calculate the threshold probability $\overline{p}$ that makes the subject S indifferent between playing A or F in response to pooling on H by the ruler R:

$$E\left(U_S\left(A,\cdot,\cdot\right);\left(NB,\cdot,\cdot,H,H\right)\right) = U_S(b_{NB})q_{\alpha l} - U_S(d_{NB})(1-q_{\alpha l})$$
$$E\left(U_S\left(F,\cdot,\cdot\right);\left(NB,\cdot,\cdot,H,H\right)\right) = U_S(a_{NB})q_{\alpha l} + U_S(c_{NB})(1-q_{\alpha l})$$

Setting $E\left(U_S\left(A,\cdot,\cdot\right);\left(NB,\cdot,\cdot,H,H\right)\right) = E\left(U_S\left(F,\cdot,\cdot\right);\left(NB,\cdot,\cdot,H,H\right)\right)$ and solving, we find:

$$q_{\alpha l}^* = \overline{p} = \frac{U_S(c_{NB}) - U_S(d_{NB})}{(U_S(b_{NB}) - U_S(a_{NB})) + (U_S(c_{NB}) - U_S(d_{NB}))}$$

We know that $0 < \overline{p} < 1$ because $U_S(c_{NB}) - U_S(d_{NB}) > 0$ and $U_S(b_{NB}) - U_S(a_{NB}) > 0$ which means that $0 < \frac{U_S(c_{NB}) - U_S(d_{NB})}{(U_S(b_{NB}) - U_S(a_{NB})) + (U_S(c_{NB}) - U_S(d_{NB}))} < 1$.

Thus, for $q_{\alpha l}^* \geq \overline{p}$, S's best response BR_S is $(A,\cdot,\cdot)$. For $q_{\alpha l}^* = \overline{p}$, BR_S is $(A,\cdot,\cdot)$ or $(F,\cdot,\cdot)$. For $q_{\alpha l}^* \leq \overline{p}$, BR_S is $(F,\cdot,\cdot)$.

A.1.1 Pooling equilibria

Lemma 1 $\{(NB,\cdot,\cdot,H,H),(A,\cdot,\cdot)\}$ *is a perfect Bayesian equilibrium of the lower game for* $p \geq \overline{p}$.

When R plays the pooling strategy $(NB,\cdot,\cdot,H,H)$ then $p = q_{\alpha l}$. BR_S is $(F,\cdot,\cdot)$ when $q_{\alpha l}^* \leq \overline{p}$. Hence, R_{ol} prefers to deviate to L as $U_R(e_{NB}) > U_R(c_{NB})$. Thus $(NB,\cdot,\cdot,H,H)$ is not part of an equilibrium for $p < \overline{p}$.

Yet BR_S is $(A, \cdot, \cdot)$ when $q^*_{\alpha l} \geq \overline{p}$. This means that R_{1l} and R_{ol} do not wish to deviate to L as $U_R(b_{NB}) > U_R(a_{NB})$ and $U_R(d_{NB}) > U_R(e_{NB})$ respectively.

Lemma 2 $\{(NB, \cdot, \cdot, L, L), (F, \cdot, \cdot)\}$ *is a perfect Bayesian equilibrium of the lower game for all* $p > 0$.

When R plays the pooling strategy $(NB, \cdot, \cdot, L, L)$ we may surmise an off-equilibrium path belief $q_{\alpha l} = 0$ such that S credibly fights F at information set α_l. Hence, R_{1l} and R_{ol} do not prefer to deviate to H as $U_R(a_{NB}) = U_R(a_{NB})$ and $U_R(e_{NB}) > U_R(c_{NB})$ respectively.

Corollary 1 $\{(NB, \cdot, \cdot, L, L), (F, \cdot, \cdot)\}$ *fails the intuitive criterion.*

The intuitive criterion corresponds with the sequential equilibrium concept rather than the perfect Bayesian equilibrium one. In this game, however, the sets of perfect Bayesian and sequential equilibria coincide.[13] Hence, we may apply the intuitive criterion.

The equilibrium $\{(NB, \cdot, \cdot, L, L), (F, \cdot, \cdot)\}$ fails the intuitive criterion as follows. S plays F in this equilibrium. Thus, R_{1l} does not receive a lower payoff for deviation to H from its equilibrium payoff to L:

$$U_{R_{1l}}(H; F) = U_{R_{1l}}(L) = U_R(a_{NB})$$

Yet R_{ol} does in fact receive a lower payoff for deviation to H from its equilibrium payoff to L:

$$U_{R_{0l}}(H; F) = U_R(c_{NB}) < U_R(e_{NB}) = U_{R_{0l}}(L; F)$$

Thus, if S believes that R_{0l} will not play H, and assigns probability 0 to R_{ol} off of the equilibrium path (i.e., $q_{\alpha l} = 1$), then S will play A as $BR_S = (A \cdot \cdot)$ when $q^*_{\alpha l} \geq \overline{p}$. Hence R_{0l} would prefer to play H, for it will receive:

$$U_{R_{0l}}(H; A) = U_R(d_{NB}) > U_R(e_{NB})$$

For this reason, the equilibrium $\{(NB, \cdot, \cdot, L, L), (F, \cdot, \cdot)\}$ is unintuitive and we do not consider it further.

Contrast this equilibrium with the pooling equilibrium $\{(NB, \cdot, \cdot, H, H), (A, \cdot, \cdot)\}$ for $p \geq \overline{p}$. S plays A in this equilibrium. R_{1l} does not receive a lower payoff for deviation to L from its equilibrium payoff to H:

$$U_{R_{1l}}(H; A) = U_R(b_{NB}) > U_R(a_{NB}) = U_{R_{1l}}(L)$$

[13] See Fudenberg and Tirole (2000), theorem 8.2, page 346. For a general discussion of the intuitive criterion, see Cho and Kreps (1987).

Nor does R_{ol} :

$$U_{R_{0l}}(H;A) = U_R(d_{NB}) > U_R(e_{NB}) = U_{R_{0l}}(L)$$

Thus, the equilibrium $\{(NB,\cdot,\cdot,H,H),(A,\cdot,\cdot)\}$ for $p \geq \overline{p}$ meets the intuitive criterion.

A.1.2 Separating Equilibria

When R plays the separating strategy $(NB,\cdot,\cdot,H,L)$ then $q_{\alpha l} = 1$. In turn, BR_S is $(A,\cdot,\cdot)$ and R_{ol} prefers to deviate to H as $U_R(d_{NB}) > U_R(e_{NB})$. Thus $(NB,\cdot,\cdot,H,L)$ is not part of an equilibrium for $0 < p < 1$. Similarly, when R plays the separating strategy $(NB,\cdot,\cdot,L,H)$ then $q_{\alpha l} = 0$. BR_S is $(F,\cdot,\cdot)$ and R_{ol} prefers to deviate to L as $U_R(e_{NB}) > U_R(c_{NB})$. Hence $(NB,\cdot,\cdot,L,H)$ is not part of an equilibrium for $0 < p < 1$.

A.1.3 Mixed Equilibria

Lemma 3 *For $p < \overline{p}$ there exists a perfect Bayesian equilibrium in mixed strategies in the lower game where the ruler R_{1l} plays H, R_{0l} mixes over H and L with $\tau_0 = \overline{\tau_0}$, and the subject mixes over A and F with $\sigma = \overline{\sigma}$.*

Let $\tau_1 = \Pr(H; X = 1)$ and $\tau_0 = \Pr(H; X = 0)$. First we ask whether R_{1l} mixes over H and L in equilibrium when $p < \overline{p}$. To start, when $0 < \tau_1 < 1$ then R_{1l} is indifferent between H and L. This indicates that S is playing F. Otherwise, R_{1l} prefers H to L as $U_R(b_{NB}) > U_R(a_{NB})$. We know that S plays F when $q^*_{\alpha l} \leq \overline{p}$.

In combination, $0 < \tau_1 < 1$ and $q^*_{\alpha l} \leq \overline{p}$ imply through Bayes' rule that $\tau_0 > 0$:

$$q_{\alpha l} = \Pr(X = 1; H) = \frac{\Pr(X = 1)\Pr(H : X = 1)}{\Pr(H)} = \frac{p \cdot \tau_1}{p \cdot \tau_1 + (1 - p) \cdot \tau_0} \leq \overline{p}$$

If $\tau_0 = 0$ then $\frac{p \cdot \tau_1}{p \cdot \tau_1 + (1-p) \cdot \tau_0} = 1 > \overline{p}$. Hence $\tau_0 > 0$.

Yet $\tau_0 > 0$ is not part of equilibrium behavior when S plays F, for R_{1l} prefers to play L rather than to mix over H and L as $U_R(c_{NB}) > U_R(e_{NB})$. This means that $\tau_0 = 0$.

However, $\tau_0 = 0$ and $q_{\alpha l} \leq \overline{p}$ are mutually inconsistent. Thus, the premise $0 < \tau_1 < 1$ is false; $\tau_1 = 1$ and R_{1l} does not mix over H and L in equilibrium.

Second we ask whether R_{0l} mixes over H and L in equilibrium when $p < \overline{p}$. R_{0l} is indifferent between H and L when $0 < \tau_0 < 1$. This indicates that S is mixing over A and F. Otherwise R_{1l} will not be indifferent between H and L since there are no payoff ties for R_{0l}; $U_R(d_{NB}) > U_R(e_{NB}) > U_R(c_{NB})$.

When R_{0l} mixes over H and L (i.e., $0 < \tau_0 < 1$) then S mixes over A and F with $\sigma = \overline{\sigma} = \frac{(U_R(e_{NB}) - U_R(d_{NB}))}{(U_R(c_{NB}) - U_R(d_{NB}))}$ to make R indifferent, where $\sigma = \Pr(F)$.[14] $0 < \overline{\sigma} < 1$ since $(U_R(c_{NB}) - U_R(d_{NB})) < (U_R(e_{NB}) - U_R(d_{NB})) < 0$, according to R's utility ranking $U_R(d_{NB}) > U_R(e_{NB}) > U_R(c_{NB})$.

S is indifferent when $\sigma = \overline{\sigma}$. This means that $U_S(F) = U_S(a_{NB})q_{\alpha l} + U_S(c_{NB})(1 - q_{\alpha l}) = U_S(A) = U_S(b_{NB})q_{\alpha l} + U_S(d_{NB})(1 - q_{\alpha l})$. Solving, we find that $q_{\alpha l} = \overline{p}$. Thus, we have internal consistency: S mixes when it is indifferent. This occurs when S believes that $q_{\alpha l} = \overline{p}$. R_{0l} must mix with the frequency that induces S's belief that $q_{\alpha l} = \overline{p}$. This implies that S is indifferent and willing to mix with $\sigma = \overline{\sigma}$. This implies in turn that R_{0l} is indifferent and willing to mix.

R_{1l}, however, strictly prefers H to L when $\sigma = \overline{\sigma}$:

$$U_R(L; X = 1) = U_R(a_{NB})$$
$$U_R(H; X = 1) = U_R(a_{NB})\sigma + U_R(b_{NB})(1 - \sigma)$$

where:

$$U_R(a_{NB}) < U_R(a_{NB})\sigma + U_R(b_{NB})(1 - \sigma), \forall \sigma \in [0, 1)$$

since $U_R(a_{NB}) < U_R(b_{NB})$.

This implies that $\tau_1 = 1$. $\tau_1 = 1$ together with $\tau_0 = \overline{\tau_0} = \frac{p(1-\overline{p})}{\overline{p}(1-p)}$ solve for $q_{\alpha l} = \overline{p}$:

$$q_{\alpha l} = \overline{p} = \frac{p \cdot \tau_1}{p \cdot \tau_1 + (1 - p) \cdot \tau_0}$$

Solving for τ_0 where $\tau_1 = 1$ means that $\tau_0 = \overline{\tau_0} = \frac{p(1-\overline{p})}{\overline{p}(1-p)}$, where $0 < \tau_0 < 1$ for $0 < p < 1$ and $0 < \overline{p} < 1$.

When $p < \overline{p}$ then $\tau_0 < \frac{\overline{p}(1-\overline{p})}{\overline{p}(1-\overline{p})} = 1$. This means that $(1 - \tau_0) < 1$.

There is no mixed strategy equilibrium, however, for $p \geq \overline{p}$. When R plays the pooling strategy $(NB, \cdot, \cdot, H, H)$ then $p = q_{\alpha l} \geq \overline{p}$. We know that BR_S is $(A, \cdot, \cdot)$ when $q^*_{\alpha l} \geq \overline{p}$. Hence, R_{1l} and R_{ol} do not wish to deviate to L; both types strictly prefer H. In turn, there is no mixing.

There is no mixed strategy equilibrium for $p = \overline{p}$ either. $\tau_0 = \frac{\overline{p}(1-\overline{p})}{\overline{p}(1-\overline{p})} = 1$ when $p = \overline{p}$. Thus, R_{ol} prefers H to any mixing.

[14] As indifference for 2 implies that $U_R(e_{NB}) = (U_R(c_{NB})\overline{\sigma} + U_R(d_{NB})(1 - \overline{\sigma}))$.

A.2 Upper subgame

As in the lower subgame, pooling, separating, and semi-separating (i.e., mixed) equilibria may arise in the upper subgame. We will now exhaust all equilibrium possibilities, eliminating most in the process.

To begin, we calculate the threshold probability $\overline{p}$ that makes the subject S indifferent between playing A or F in response to pooling on H by the ruler R. Setting $E\left(U_S\left(\cdot, A, \cdot\right); \left(B, H, H, \cdot, \cdot\right)\right) = E\left(U_S\left(\cdot, F, \cdot\right); \left(B, H, H, \cdot, \cdot\right)\right)$ and solving, we find that $q^*_{\alpha u} = \overline{p} = \frac{U_S(c_B) - U_S(d_B)}{(U_S(b_B) - U_S(a_B)) + (U_S(c_B) - U_S(d_B))} = \frac{U_S(c_{NB}) - U_S(d_{NB})}{(U_S(b_{NB}) - U_S(a_{NB})) + (U_S(c_{NB}) - U_S(d_{NB}))}$ for the upper subgame as well, where $0 < \overline{p} < 1$. Equality between the upper and lower subgames holds because we assume that $U_S(a_B) = (U_S(a_{NB}) - I_S)$, $U_S(b_B) = (U_S(b_{NB}) - I_S)$, $U_S(c_B) = (U_S(c_{NB}) - I_S)$, and $U_S(d_B) = (U_S(d_{NB}) - I_S)$ (see section 2.4.2).

Thus, for $q^*_{\alpha u} \geq \overline{p}$, S's best response BR_S is $(\cdot, A, \cdot)$. For $q^*_{\alpha u} = \overline{p}$, BR_S is $(\cdot, A, \cdot)$ or $(\cdot, F, \cdot)$. For $q^*_{\alpha u} \leq \overline{p}$, BR_S is $(\cdot, F, \cdot)$.

A.2.1 Pooling equilibria

Lemma 4 $\{(B, H, H, \cdot, \cdot), (\cdot, A, \cdot)\}$ *is a perfect Bayesian equilibrium of the upper game for* $p \geq \overline{p}$.

When R plays the pooling strategy $(B, H, H, \cdot, \cdot)$ then $p = q_{\alpha u}$. BR_S is $(\cdot, F, \cdot)$ when $q^*_{\alpha u} \leq \overline{p}$. Hence, R_{ou} prefers to deviate to L as $U_R\left(e_B\right) > U_R\left(c_B\right)$. Thus $(B, H, H, \cdot, \cdot)$ is not part of an equilibrium for $p < \overline{p}$.

Yet BR_S is $(\cdot, A, \cdot)$ when $q^*_{\alpha u} \geq \overline{p}$. This means that R_{1u} and R_{ou} do not prefer to deviate to L as $U_R\left(b_B\right) > U_R(a_B)$ and $U_R\left(d_B\right) > U_R\left(e_B\right)$ respectively.

Lemma 5 $\{(B, L, L, \cdot, \cdot), (\cdot, F, \cdot)\}$ *is a perfect Bayesian equilibrium of the upper game for all* $p > 0$.

When R plays the pooling strategy $(B, L, L, \cdot, \cdot)$ we may surmise an off-equilibrium path belief $q_{\alpha u} = 0$ such that R credibly fights F at information set α. Hence R_{1u} and R_{ou} do not prefer to deviate to H as $U_R(a_B) = U_R(a_B)$ and $U_R\left(e_B\right) > U_R(c_B)$ respectively.

Corollary 2 $\{(B, L, L, \cdot, \cdot), (\cdot, F, \cdot)\}$ *fails the intuitive criterion.*

As for the lower game, $\{(B, L, L, \cdot, \cdot), (\cdot, F, \cdot)\}$ is unintuitive (see section A.1.1). Thus, it is eliminated from further consideration.

When R plays the pooling strategy $(B, CS, CS, \cdot, \cdot)$ then $p = q_\gamma$. BR_S is $(\cdot, \cdot, A)$ for all $p > 0$ as $U_S\left(f_B\right) > U_S\left(a_B\right)$ and $U_S\left(a_B\right) = U_S\left(a_B\right)$ respectively. Hence, R_{ou} prefers to deviate to L as $U_R(e_B) > U_R(a_B)$. Thus $(B, CS, CS, \cdot, \cdot)$ is not part of an equilibrium for $p > 0$.

A.2.2 Separating Equilibria

Lemma 6 $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ *is a perfect Bayesian equilibrium of the upper game for* $0 < p < 1$.

When R plays the separating strategy $(B, H, L, \cdot, \cdot)$ then $q_{\alpha u} = 1$. In turn, BR_S is $(\cdot, A, \cdot)$ and R_{ou} prefers to deviate to H as $U_R(d_B) > U_R(e_B)$. Thus $(B, H, L, \cdot, \cdot)$ is not part of an equilibrium for $0 < p < 1$. Similarly, when R plays the separating strategy $(B, L, H, \cdot, \cdot)$ then $q_{\alpha u} = 0$. BR_S is $(\cdot, F, \cdot)$ and R_{ou} prefers to deviate to L as $U_R(e_B) > U_R(c_B)$. Hence $(B, L, H, \cdot, \cdot)$ is not part of an equilibrium for $0 < p < 1$.

When R plays the separating strategy $(B, CS, H, \cdot, \cdot)$ then $q_\gamma = 1$ and $q_{\alpha u} = 0$. BR_S is $(\cdot, F, A)$. Hence R_{ou} prefers to deviate to L as $U_R(e_B) > U_R(c_B)$. Thus $(B, CS, H, \cdot, \cdot)$ is not part of an equilibrium for all $0 < p < 1$.

When R plays the separating strategy $(B, CS, L, \cdot, \cdot)$ then $q_{\alpha u} = 0$ and $q_\gamma = 1$. BR_S is $(\cdot, F, A)$. Hence R_{1u} and R_{ou} do not prefer to deviate to H as $U_R(f_B) > U_R(a_B)$ and $U_R(e_B) > U_R(c_B)$. This implies:

When R plays the separating strategy $(B, H, CS, \cdot, \cdot)$ then $q_{\alpha u} = 1$ and $q_\gamma = 0$. BR_S is $(\cdot, A, A)$. Hence R_{ou} prefers to deviate to L as $U_R(e_B) > U_R(a_B)$. Thus $(B, H, CS, \cdot, \cdot)$ is not part of an equilibrium for $0 < p < 1$. Similarly, when R plays the separating strategy $(B, L, CS, \cdot, \cdot)$ then $q_{\alpha u} = 0$ and $q_\gamma = 0$. BR_S is $(\cdot, F, A)$. Again R_{ou} prefers to deviate to L as $U_R(e_B) > U_R(c_B)$. Thus $(B, L, CS, \cdot, \cdot)$ is not part of an equilibrium for $0 < p < 1$.

A.2.3 Mixed Equilibria

There is no mixed strategy equilibrium when $p < \overline{p}$. We know that R_{1u} does not mix over CS, H, and L in equilibrium. The logic resembles that of the lower subgame (see section A.1.3). For R_{ou} to mix, it must be indifferent between CS, H, and L. For $p < \overline{p}$, R_{1u} prefers CS; there is no longer pooling on H as $U_R(f_B) > U_R(a_B)$. This implies that $q_{\alpha u} = 0$. BR_S is $(\cdot, F, \cdot)$. Hence R_{ou} is no longer indifferent between CS, H, and L. In fact, R_{ou} prefers L to H as $U_R(e_B) > U_R(c_B)$ and L to CS as $U_R(e_B) > U_R(a_B)$. This means that there is no mixed strategy equilibrium for $p < \overline{p}$.

Similarly, there is no mixed strategy equilibrium for $p \geq \overline{p}$. When R plays the pooling strategy $(B, H, H, \cdot, \cdot)$ then $q_{\alpha u} = 1 \geq \overline{p}$. BR_S is $(\cdot, A, \cdot)$ when $q^*_{\alpha u} \geq \overline{p}$. Hence R_{1u} and R_{ou} do not prefer to deviate to L or CS; both types strictly prefer H as $U_R(b_B) > U_R(a_B)$

and $U_R(d_B) > U_R(e_B)$ and as $U_R(b_B) > U_R(f_B)$ and $U_R(d_B) > U_R(a_B)$ respectively. Thus there is no mixed strategy equilibrium for $p \geq \overline{p}$.

A.3 Game in total

We now compare lower and upper subgame equilibria for $p \geq \overline{p}$ and $p \leq \overline{p}$ to find equilibria for the game in total. Let us recap. For the lower subgame, there are two equilibria: the predatory equilibrium $\{(NB, \cdot, \cdot, H, H), (A, \cdot, \cdot)\}$ for $p \geq \overline{p}$ and the mixed equilibrium for $p < \overline{p}$. There are also two equilibria for the upper subgame; the predatory equilibrium $\{(B, H, H, \cdot, \cdot), (\cdot, A, \cdot)\}$ for $p \geq \overline{p}$ and the credible-signal equilibrium $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ for $0 \leq p < 1$. Thus before we can compare upper and lower subgame outcomes we must narrow down the upper subgame equilibria.

Lemma 7 $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ *is the unique perfect Bayesian equilibrium of the upper game for* $p < \overline{p}$.

For $p < \overline{p}$ the credible-signal equilibrium $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ is unique.

Lemma 8 $\{(B, H, H, \cdot, \cdot), (\cdot, A, \cdot)\}$ *is the unique perfect Bayesian equilibrium of the upper game for* $p \geq \overline{p}$.

We have two possibilities, however, for $p \geq \overline{p}$: the credible-signal equilibrium and the predatory equilibrium. Hence we must determine which outcome R prefers:

$$E(U_R((B, CS, L, \cdot, \cdot); (\cdot, F, A))) = U_R(f_B)\, p + U_R(e_B)(1-p)$$
$$E(U_R((B, H, H, \cdot, \cdot); (\cdot, A, \cdot))) = U_R(b_B)\, p + U_R(d_B)(1-p)$$

We know that $U_R(b_B) > U_R(f_B), U_R(e_B)$ and $U_R(d_B) > U_R(f_B), U_R(e_B)$.

Thus $E(U_R((B, H, H, \cdot, \cdot); (\cdot, A, \cdot))) > E(U_R((B, CS, L, \cdot, \cdot); (\cdot, F, A)))$ and R prefers $s_R = (B, H, H, \cdot, \cdot)$ to $s_R = (B, CS, L, \cdot, \cdot)$.

We are ready to compare lower and upper subgame outcomes for the game in total. For exposition purposes, we cover the proofs in the following order: Proposition 2, Proposition 4, Remark 1, Remark 2, and Propositions 1 and 3.

Proof of Proposition 2. When $p \geq \overline{p}$, R plays the upper subgame equilibrium strategy $(B, H, H, \cdot, \cdot)$ in response to S's equilibrium strategy $(A, \cdot, \cdot)$ and the lower subgame equilibrium strategy $(NB, \cdot, \cdot, H, H)$ in response to S's equilibrium strategy $(\cdot, A, \cdot)$:

$$\begin{aligned} E(U_R((B, H, H, \cdot, \cdot), (\cdot, A, \cdot))) &= U_R(b_B)\, p + U_R(d_B)(1-p) \\ E(U_R((NB, \cdot, \cdot, H, H); (A, \cdot, \cdot))) &= U_R(b_{NB}) p + U_R(d_{NB})(1-p) \end{aligned}$$

We know that $U_R(b_{NB}) > U_R(b_B)$ and $U_R(d_{NB}) > U_R(d_B)$.

Thus $E(U_R((NB,\cdot,\cdot,H,H);(A,\cdot,\cdot))) > E(U_R((B,H,H,\cdot,\cdot);(\cdot,A,\cdot)))$ and R prefers $s_R = (NB,\cdot,\cdot,H,H)$ to $s_R = (B,H,H,\cdot,\cdot)$.

Proof of Proposition 4. When $p < \overline{p}$, R plays the upper subgame equilibrium strategy $(B, CS, L,\cdot,\cdot)$ in response to S's equilibrium strategy $(\cdot, F, A)$ and the lower subgame mixed equilibrium strategy in response to S's mixed equilibrium strategy:

$$E(U_R((B,CS,L,\cdot,\cdot);(\cdot,F,A))) = U_R(f_B)\,p + U_R(e_B)\,(1-p)$$

$$E(U_R(NB,\text{ mix})) = (U_R(a_{NB})\overline{\sigma} + U_R(b_{NB})\,(1-\overline{\sigma}))\,p + \begin{pmatrix} U_R(c_{NB})\overline{\sigma} \\ +U_R(d_{NB})\,(1-\overline{\sigma}) \end{pmatrix}(1-p)$$

Recognizing from section A.1.3 that $U_R(e_{NB}) = (U_R(c_{NB})\overline{\sigma} + U_R(d_{NB})\,(1-\overline{\sigma}))$, we set $E(U_R((B,CS,L,\cdot,\cdot);(\cdot,F,A))) = E(U_R(NB,\text{ mix}))$ to find:

$$p = \widetilde{p} = \frac{I_R}{U_R(e_{NB}) - U_R(a_{NB})\overline{\sigma} - U_R(b_{NB})(1-\overline{\sigma})}$$

where $U_R(e_B) = U_R(f_B) = (U_R(e_{NB}) - I_R)$ (see section 2.4.2).

Substituting for $\overline{\sigma} = \frac{(U_R(e_{NB})-U_R(d_{NB}))}{(U_R(c_{NB})-U_R(d_{NB}))}$, we find:

$$\widetilde{p} = \frac{I_R(U_R(c_{NB}) - U_R(d_{NB}))}{U_R(d_{NB})(U_R(a_{NB}) - U_R(e_{NB})) + U_R(c_{NB})\,(U_R(e_{NB}) - U_R(b_{NB})) + U_R(e_{NB})(U_R(b_{NB}) - U_R(a_{NB}))}$$
$$= \frac{I_R(U_R(c_{NB}) - U_R(d_{NB}))}{\phi}$$

where:

$$\phi = U_R(d_{NB})(U_R(a_{NB}) - U_R(e_{NB})) + U_R(c_{NB})\,(U_R(e_{NB}) - U_R(b_{NB})) + U_R(e_{NB})(U_R(b_{NB}) - U_R(a_{NB}))$$

We know that $(U_R(c_{NB}) - U_R(d_{NB})) < 0$. This implies that $\phi < 0$ in order for $0 < \widetilde{p} < 1$.

To determine whether $\widetilde{p} > \overline{p}$ or $\widetilde{p} < \overline{p}$, we set $\overline{p} = \widetilde{p}$ and solve for $I_R = \overline{I}_R$:

$$\overline{p} = \frac{U_S(c_{NB}) - U_S(d_{NB})}{(U_S(b_{NB}) - U_S(a_{NB})) + (U_S(c_{NB}) - U_S(d_{NB}))} = \widetilde{p} = \frac{I_R(U_R(c_{NB}) - U_R(d_{NB}))}{\phi}$$

This implies that:

$$I_R = \overline{I}_R = \left(\frac{U_S(c_{NB}) - U_S(d_{NB})}{(U_R(c_{NB}) - U_R(d_{NB}))}\right)\left(\frac{\phi}{(U_S(b_{NB}) - U_S(a_{NB})) + (U_S(c_{NB}) - U_S(d_{NB}))}\right)$$

If $I_R > \overline{I}_R$ then $\widetilde{p} = \frac{I_R(U_R(c_{NB})-U_R(d_{NB}))}{\phi}$ increases. Thus $\widetilde{p} > \overline{p}$ and R prefers the mixed equilibrium strategy to the credible-signal equilibrium strategy as $E(U_R(NB, \text{mix})) > E(U_R((B, CS, L, \cdot, \cdot)\,;(\cdot, F, A)))$. That is, R does not prefer $s_R = (B, CS, L, \cdot, \cdot)$ over the mixed strategy unless $p > \widetilde{p} > \overline{p}$. And R prefers $s_R = (NB, \cdot, \cdot, H, H)$ over $s_R = (B, CS, L, \cdot, \cdot)$ when $p \geq \overline{p}$. Thus, R does not play $s_R = (B, CS, L, \cdot, \cdot)$ under this payoff structure.

Proof of Remark 1. $I_R = 0$ means that R bears none of the cost of building the credible-signal institution. This implies that $\widetilde{p} = \frac{I_R(U_R(c_{NB})-U_R(d_{NB}))}{\phi} = 0$. See the proof of Proposition 4 for the derivation of $\widetilde{p}$.

That is, R does not play the mixed equilibrium strategy for $p < \overline{p}$ when $I_R = 0$. We confirm this as follows:

$$E(U_R((B, CS, L, \cdot, \cdot)\,;(\cdot, F, A))) = U_R(f_B)p + U_R(e_B)(1-p)$$

$$E(U_R(NB, \text{mix})) = (U_R(a_{NB})\overline{\sigma} + U_R(b_{NB})\,(1-\overline{\sigma}))\,p + \begin{pmatrix} U_R(c_{NB})\overline{\sigma} + \\ U_R(d_{NB})\,(1-\overline{\sigma}) \end{pmatrix}(1-p)$$

Recognizing from section A.1.3 that $U_R(e_{NB}) = (U_R(c_{NB})\overline{\sigma} + U_R(d_{NB})\,(1-\overline{\sigma}))$, we show that $U_R(f_B) > U_R(a_{NB})\overline{\sigma} + U_R(b_{NB})\,(1-\overline{\sigma})$. Substituting for $\overline{\sigma} = \frac{(U_R(e_{NB})-U_R(d_{NB}))}{(U_R(c_{NB})-U_R(d_{NB}))}$, we have:

$$U_R(f_B) > U_R(a_{NB})\left(\frac{(U_R(e_{NB}) - U_R(d_{NB}))}{(U_R(c_{NB}) - U_R(d_{NB}))}\right) + U_R(b_{NB})\left(1 - \frac{(U_R(e_{NB}) - U_R(d_{NB}))}{(U_R(c_{NB}) - U_R(d_{NB}))}\right)$$

Recall that $U_R(e_B) = U_R(f_B) = (U_R(e_{NB}) - I_R)$ (see section 2.4.2). Hence, $U_R(f_B) = U_R(e_{NB})$ when $I_R = 0$.

Replacing $U_R(f_B)$ with $U_R(e_{NB})$ and rearranging, we find:

$$\begin{aligned} U_R(e_{NB})(U_R(c_{NB}) - U_R(d_{NB})) \quad < \quad & U_R(b_{NB})U_R(c_{NB}) - U_R(a_{NB})U_R(d_{NB}) \\ & + U_R(a_{NB})U_R(e_{NB}) - U_R(b_{NB})U_R(e_{NB}) \end{aligned}$$

Recall that R has the following utility rankings when $I_R = 0$ (see section 2.4.2):

$$U_R(d_{NB}) > U_R(b_{NB}) > U_R(e_{NB}) > U_R(c_{NB}) > U_R(a_{NB})$$

Normalizing $U_R(d_{NB}) = 1$ and $U_R(a_{NB}) = 0$, we are able to simplify:

$$U_R(b_{NB})U_R(c_{NB}) - U_R(b_{NB})U_R(e_{NB}) - U_R(e_{NB})U_R(c_{NB}) + U_R(e_{NB}) > 0$$

This implies:

$$U_R(c_{NB})\left(U_R(b_{NB}) - U_R(e_{NB})\right) + U_R(e_{NB})\left(1 - U_R(b_{NB})\right) > 0$$

which holds since $(U_R(b_{NB}) - U_R(e_{NB})) > 0$, $(1 - U_R(b_{NB})) > 0$, and $U_R(c_{NB}), U_R(e_{NB}) > 0$, according to our normalization.[15]

Remark 2 *So long as the ruler bears no more than $I_R \leq \overline{I}_R$ of the cost of building the credible-signal institution, there exists a perfect Bayesian equilibrium $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$ for $p < \overline{p}$, where I_R, $\overline{I}_R > 0$.*

If on the other hand $I_R < \overline{I}_R$, then $\widetilde{p} = \frac{I_R(U_R(c_{NB}) - U_R(d_{NB}))}{\phi}$ decreases. Thus $\widetilde{p} < \overline{p}$ and the credible-signal equilibrium strategy $(B, CS, L, \cdot, \cdot)$ becomes optimal for at least some $p < \overline{p}$ values.[16] See the proof of Proposition 4 for the derivation of $\widetilde{p}$.

Proof of Propositions 1 and 3. We refine remark 2 by introducing an intermediate cost burden $I_R = \widehat{I_R}$, where $0 < \widehat{I_R} < \overline{I}_R$, such that R prefers the mixed equilibrium strategy for $p < \widehat{p} < \overline{p}$ and the credible-signal equilibrium strategy for $\widehat{p} \leq p < \overline{p}$. To find the $p = \widehat{p}$ value that corresponds to this intermediate value $\widehat{I_R}$:

$$E(U_R(NB, \text{mix})) = \left(U_R(a_{NB})\overline{\sigma} + U_R(b_{NB})\left(1 - \overline{\sigma}\right)\right)p + \begin{pmatrix} U_R(c_{NB})\overline{\sigma} + \\ U_R(d_{NB})\left(1 - \overline{\sigma}\right) \end{pmatrix}(1 - p)$$

And:

$$\begin{aligned} E(U_R((B, CS, L, \cdot, \cdot); (\cdot, F, A))) &= U_R(f_B)p + U_R(e_B)(1 - p) \\ &= \left(U_R(e_{NB}) - \widehat{I_R}\right) \end{aligned}$$

[15] Thanks to Joon Song for his help with this proof.

[16] Assume that, so long as $p < \overline{p}$, R plays the credible-signal strategy $(B, CS, L, \cdot, \cdot)$ over the mixed strategy at the point of indifference $I_R = \overline{I}_R$.

where $U_R(e_B) = U_R(f_B) = \left(U_R(e_{NB}) - \widehat{I_R}\right)$ (see section 2.4.2).

Recognizing again from section A.1.3 that $U_R(e_{NB}) = (U_R(c_{NB})\overline{\sigma} + U_R(d_{NB})(1-\overline{\sigma}))$, we set $E(U_R((B, CS, L, \cdot, \cdot); (\cdot, F, A))) = E(U_R(NB, \text{mix}))$ to find:

$$I_R = \widehat{I_R} = p(U_R(e_{NB}) - U_R(a_{NB})\overline{\sigma} - U_R(b_{NB})(1-\overline{\sigma}))$$

This implies:

$$p = \widehat{p} = \frac{\widehat{I_R}}{U_R(e_{NB}) - U_R(a_{NB})\overline{\sigma} - U_R(b_{NB})(1-\overline{\sigma})}$$

Substituting for $\overline{\sigma} = \frac{(U_R(e_{NB}) - U_R(d_{NB}))}{(U_R(c_{NB}) - U_R(d_{NB}))}$, we find:

$$\widehat{p} = \frac{\widehat{I_R}(U_R(c_{NB}) - U_R(d_{NB}))}{\phi}$$

where again:

$$\phi = U_R(d_{NB})(U_R(a_{NB}) - U_R(e_{NB})) + U_R(c_{NB})(U_R(e_{NB}) - U_R(b_{NB})) + U_R(e_{NB})(U_R(b_{NB}) - U_R(a_{NB}))$$

Again, we know that $(U_R(c_{NB}) - U_R(d_{NB})) < 0$, which implies that $\phi < 0$ in order for $0 < \widehat{p} < 1$.

We now have the exact range of p values (i.e., given $p < \overline{p}$) over which R prefers to play the credible-signal equilibrium strategy and the mixed equilibrium strategy. Assume that, at the point of indifference $p = \widehat{p}$, R plays the credible signal strategy $\{(B, CS, L, \cdot, \cdot), (\cdot, F, A)\}$.

A.4 Repeated Game

Lemma 9 *There exists a discount factor $\delta(p)$ such that for $\delta(p) \geq \overline{\delta}(p)$ the strategy profile for the ruler and the subject with equilibrium behavior $\{(NB, \cdot, \cdot, H, L), (A, \cdot, \cdot)\}$ is a subgame perfect equilibrium of the infinitely repeated lower subgame.*

We wish to maintain a separating (i.e., "honesty") equilibrium for the infinitely repeated game where R is the long-run player and S is the short-run player. In the lower subgame, this takes the form $\{(NB, \cdot, \cdot, H, L), (A, \cdot, \cdot)\}$, where the maintenance payoff is:

$$E(U_R((NB, \cdot, \cdot, H, L); (A, \cdot, \cdot))) = U_R(b_{NB})p + U_R(e_{NB})(1-p)$$

The deviation payoff is:

$$E(U_R((NB \cdot \cdot HH);(A,\cdot,\cdot))) = U_R(b_{NB})p + U_R(d_{NB})(1-p)$$

The reversion payoff is:

$$E(U_R((NB,\cdot,\cdot,L,L)\,;(F,\cdot,\cdot)) = U_R(a_{NB})p + U_R(e_{NB})(1-p)$$

Hence, for discount factor $\delta\,(p)$, we must have:

$$U_R(b_{NB})p+U_R(e_{NB})(1-p) \geq (1-\delta)\left(U_R(b_{NB})p + U_R(d_{NB})(1-p)\right)+\delta\left(U_R(a_{NB})p + U_R(e_{NB})(1-p)\right)$$

Solving for the threshold discount factor $\overline{\delta}\,(p)$:

$$\delta\,(p) \geq \overline{\delta}\,(p) = \frac{(1-p)\left(U_R(d_{NB}) - U_R(e_{NB})\right)}{p\left(U_R(b_{NB}) - U_R(a_{NB})\right) + (1-p)\left(U_R(d_{NB}) - U_R(e_{NB})\right)}$$

where $0 \leq \overline{\delta}\,(p) \leq 1$ since $0 < (U_R(d_{NB}) - U_R(e_{NB})) < (U_R(b_{NB}) - U_R(a_{NB}))+(U_R(d_{NB}) - U_R(e_{NB}))$ and $0 \leq p \leq 1$. This implies the following strategy profile.

- R: Play $(NB,\cdot,\cdot,H,L)$ in the first stage and continue to do so as long as the history of play is $\{(NB,\cdot,\cdot,H,L)\,,(A,\cdot,\cdot)\}$. Otherwise, play $(NB,\cdot,\cdot,L,L)$.
- S: Play $(A,\cdot,\cdot)$ in the first stage and continue to do so as long as the history of play is $\{(NB,\cdot,\cdot,H,L)\,,(A,\cdot,\cdot)\}$. Otherwise, play $(F,\cdot,\cdot)$.

Lemma 10 *There exists a discount factor $\delta(p)$ such that for $\delta(p) \geq \overline{\delta}(p)$ the strategy profile for the ruler and the subject with equilibrium behavior $\{(B,H,L,\cdot,\cdot)\,,(\cdot,A,\cdot)\}$ is a subgame perfect equilibrium of the infinitely repeated upper subgame.*

In the upper subgame, the honesty equilibrium takes the form $\{(B,H,L,\cdot,\cdot)\,,(\cdot,A,\cdot)\}$, where the maintenance payoff is:

$$E(U_R((B,H,L,\cdot,\cdot)\,,(\cdot,A,\cdot))) = U_R\,(b_B)\,p + U_R\,(e_B)\,(1-p)$$

The deviation payoff is:

$$E(U_R((B,H,H,\cdot,\cdot)\,,(\cdot,A,\cdot))) = U_R\,(b_B)\,p + U_R(d_B)(1-p)$$

The reversion payoff is:

$$E(U_R((B,L,L,\cdot,\cdot),(\cdot,F,\cdot)) = U_R(a_B)p + U_R(e_B)(1-p)$$

Hence, for discount factor $\delta(p)$, we must have:

$$U_R(b_B)p + U_R(e_B)(1-p) \geq (1-\delta)(U_R(b_B)p + U_R(d_B)(1-p) + \delta(U_R(a_B)p + U_R(e_B)(1-p))$$

Solving for the threshold discount factor $\overline{\delta}(p)$:

$$\delta(p) \geq \frac{(1-p)(U_R(d_B) - U_R(e_B))}{p(U_R(b_B) - U_R(a_B)) + (1-p)(U_R(d_B) - U_R(e_B))}$$

Simplifying:

$$\delta(p) \geq \overline{\delta}(p) = \frac{(1-p)(U_R(d_{NB}) - U_R(e_{NB}))}{p(U_R(b_{NB}) - U_R(a_{NB})) + (1-p)(U_R(d_{NB}) - U_R(e_{NB}))}$$

since $U_R(a_B) = (U_R(a_{NB}) - I_R)$, $U_R(b_B) = (U_R(b_{NB}) - I_R)$, $U_R(d_B) = (U_R(d_{NB}) - I_R)$, and $U_R(e_B) = (U_R(e_{NB}) - I_R)$ (see section 2.4.2). Hence, the threshold discount factor $\overline{\delta}(p)$ is the same for both the lower and upper subgames. This implies the following strategy profile:

- R: Play $(B,H,L,\cdot,\cdot)$ in the first stage and continue to do so as long as the history of play is $\{(BHL),(\cdot,A,\cdot)\}$. Otherwise, play $(B,CS,L,\cdot,\cdot)$.
- S: Play $(\cdot,A,\cdot)$ in the first stage and continue to do so as long as the history of play is $\{(B,H,L,\cdot,\cdot),(\cdot,A,\cdot)\}$. Otherwise, play $(\cdot,F,\cdot)$.

Proof of Proposition 5. To find the honesty equilibrium for the infinitely repeated game in total, we must compare equilibria for the upper and lower subgames:

$$\begin{aligned} E(U_R((B,H,L,\cdot,\cdot),(\cdot,A,\cdot))) &= U_R(b_B)p + U_R(e_B)(1-p) \\ E(U_R((NB,\cdot,\cdot,H,L);(A,\cdot,\cdot))) &= U_R(b_{NB})p + U_R(e_{NB})(1-p) \end{aligned}$$

R prefers $s_R = (NB,\cdot,\cdot,H,L)$ to $s_R = (B,H,L,\cdot,\cdot)$ so long as $I_R > 0$, since $U_R(b_{NB}) > U_R(b_B)$ and $U_R(e_{NB}) > U_R(e_B)$. Assume that R continues to play $s_R = (NB,\cdot,\cdot,H,L)$ at the point of indifference $I_R = 0$.

Note Corresponding to Figure 6. The first derivative of $\overline{\delta}(p)$ with respect to ∂p is:

$$\frac{\partial\overline{\delta}(p)}{\partial p} = \frac{U_R(d_{NB})\left(U_R(e_{NB}) - U_R(d_{NB})\right)}{\Psi^2} > 0$$

where $\Psi = p\left(U_R(b_{NB}) - U_R(a_{NB})\right) + (1-p)\left(U_R(d_{NB}) - U_R(e_{NB})\right)$.

The second derivative of $\overline{\delta}(p)$ with respect to ∂p is:

$$\frac{\partial^2\overline{\delta}(p)}{\partial p^2} = \frac{2\Psi(U_R(d_{NB})\left(U_R(e_{NB}) - U_R(d_{NB})\right)\left(\left(U_R(b_{NB}) + U_R(e_{NB})\right) - \left(U_R(a_{NB}) + U_R(d_{NB})\right)\right))}{\Psi^4 e_{NB}} > 0$$

so long as $\left(U_R(b_{NB}) + U_R(e_{NB})\right) > \left(U_R(a_{NB}) + U_R(d_{NB})\right)$.

References

[1] Acemoglu, D (2005). "Politics and Economics in Weak and Strong States." *NBER Working Paper 11275.*

[2] Acemoglu, D., S. Johnson, and J. Robinson (2001). "Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5), 1369-1401.

[3] Acemoglu, D. and J. Robinson (2000) "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics*, 115 (4), 1167-1199.

[4] Aghion, P., A. Alesina, and F. Trebbi (2004) "Endogenous Political Institutions." *Quarterly Journal of Economics*, 119 (2), 565-605.

[5] Cho, I.K. and D. Kreps (1987). "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics*, 102 (2), 179-221.

[6] Fudenberg, D. and J. Tirole (2000). *Game Theory.* Cambridge (MA): MIT Press.

[7] Green, E. (1993) "On the Emergence of Parliamentary Government: The Role of Private Information." *Quarterly Review, Federal Reserve Bank of Minneapolis*, 17 (1), 2-16.

[8] Lizzeri, A. and N. Persico (2003) "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform.'" *Journal of Economic Literature*, 119 (2), 565-605.

[9] McGuire, M. and M. Olson (1996) "The Economics of Autocracy and Majority Rule: The Invisible Hand and the Use of Force." *Journal of Economic Literature*, 34 (1), 72-96.

[10] North, D. (1981). *Structure and Change in Economic History.* New York: Norton.

[11] North, D. and B Weingast (1989) "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in 17th Century England." *Journal of Economic History*, 49 (4), 803-832.

[12] Olson, M. (1982). *The Rise and Decline of Nations: Economic Growth, Stagflation, and Economic Rigidities.* New Haven: Yale University Press.

[13] Stasavage, David (2003). *Public Debt and the Birth of the Democratic State: France and Great Britain, 1688-1789*. Cambridge: Cambridge University Press.

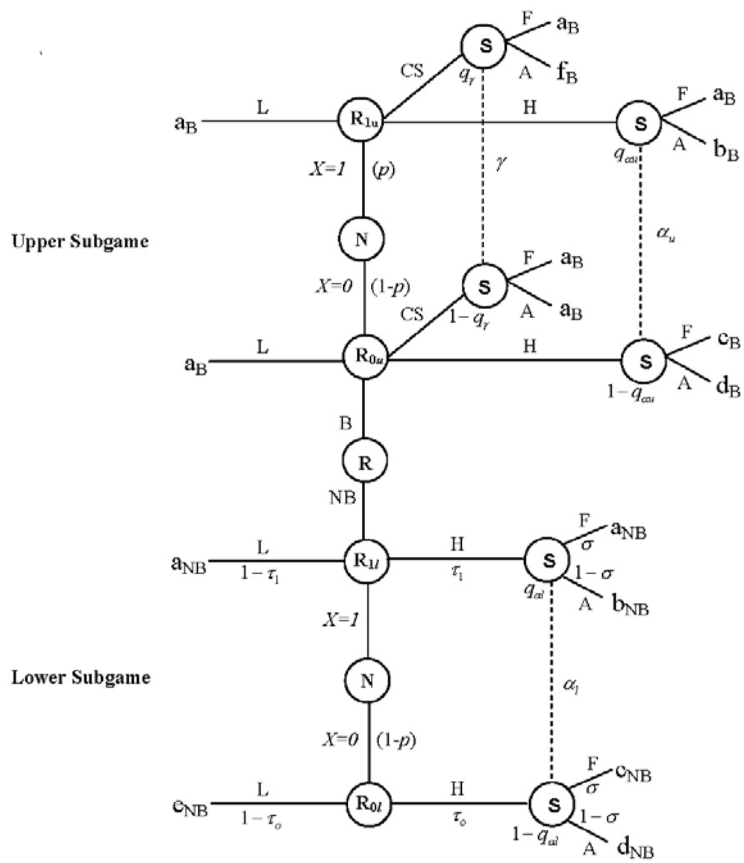

Figure 1: Game Tree with General Notation

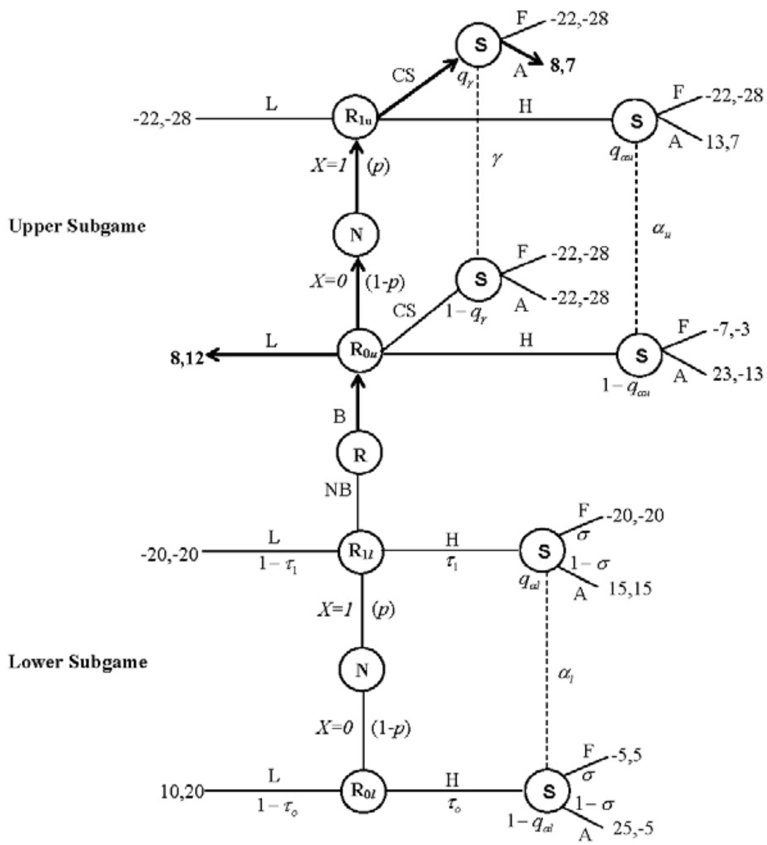

Figure 2: Example of Credible-Signal Equilibrium for $\widehat{p} \leq p < \overline{p}$ with $I_R = 2$, $I_S = 8$, and $I = I_R + I_S = 10$

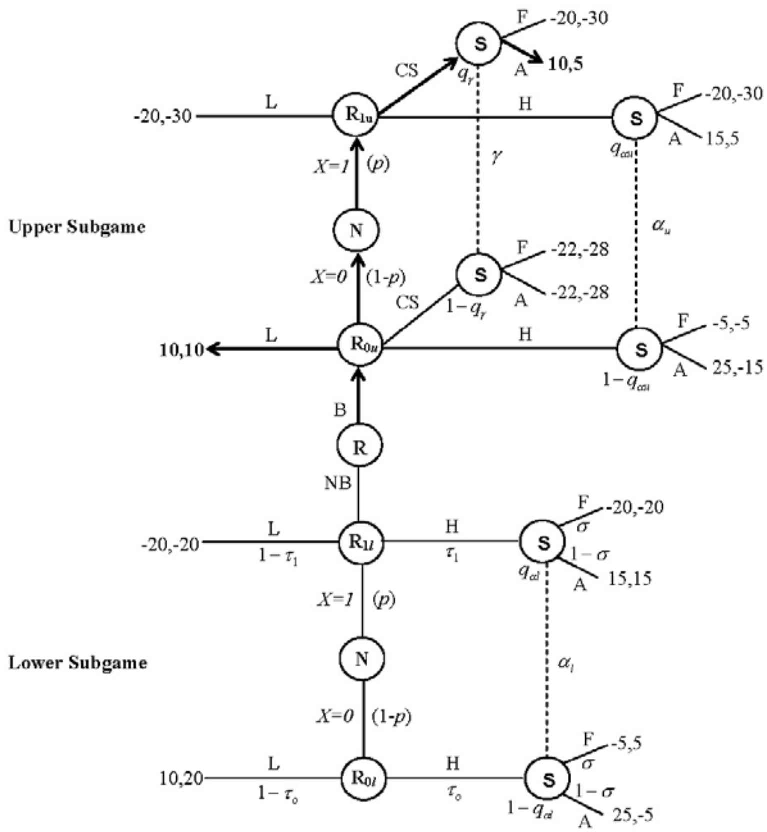

Figure 3: Example of Credible-Signal Equilibrium for $p < \overline{p}$ with $I_R = 0$ and $I_S = I = 10$

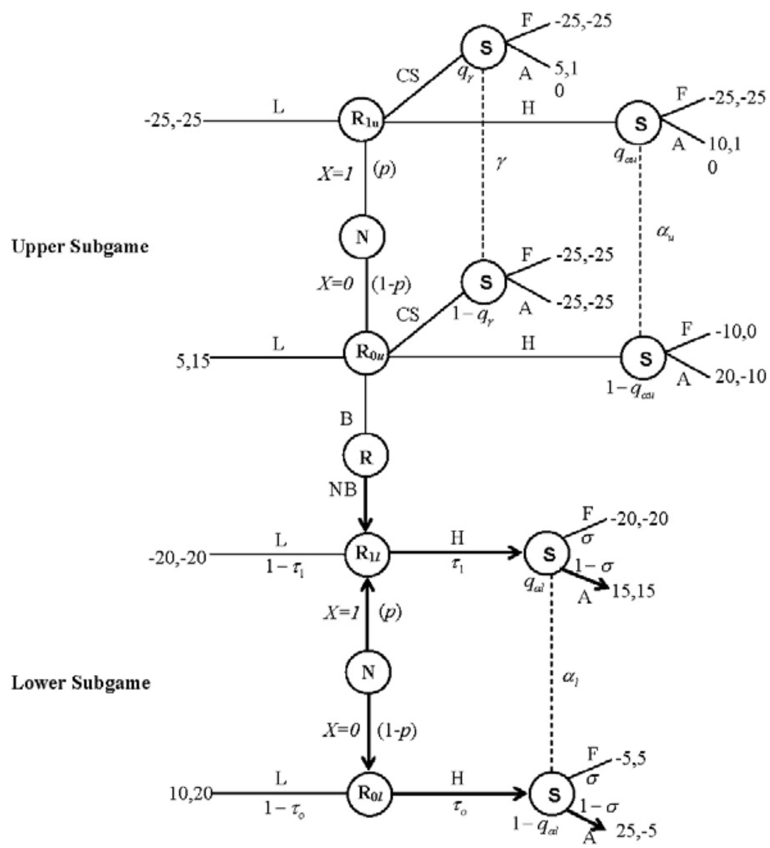

Figure 4: Example of Predatory Equilibrium for $p \geq \overline{p}$ with $I_R = 0$ and $I_S = I = 10$

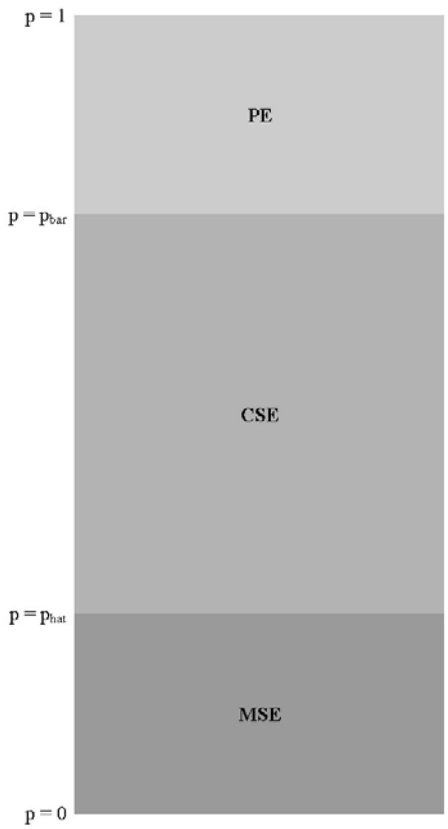

Figure 5: Predatory (PE), Credible-Signal (CSE), and Mixed Strategy (MSE) equilibria for $0 < I_R \leq \overline{I}_R$ and $\delta(p) = 0$

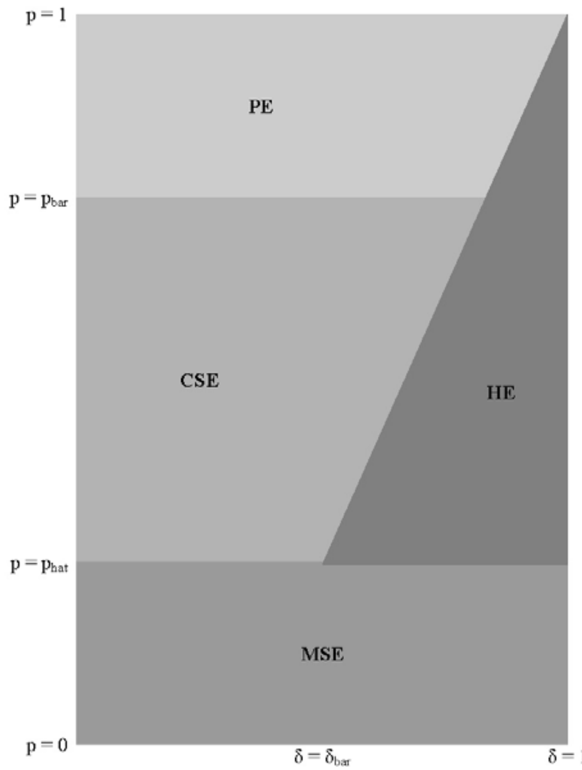

Figure 6: Approximation of Predatory (PE), Credible-Signal (CSE), Mixed Strategy (MSE), and Honesty (HE) equilibria for $0 < I_R \leq \overline{I}_R$ and $0 \leq \overline{\delta}\,(p) \leq 1$