Sanctions as Instruments of Regime Change
Online Appendix

Andrew J. Coe and Stephanie N. Kang

Contents

1 Proofs of Propositions

Lemma 1. *Let $\overline{\sigma}$ be such that $c(\overline{\sigma}) = \phi$. If S threatens sanctions of a severity higher than $\overline{\sigma}$, she will never impose them, and so R and T will ignore her demand. If S threatens sanctions of a severity no more than $\overline{\sigma}$, she will impose them if her demand is not met.*

Proof. Consider any terminal subgame of the game, in which S decides whether to impose sanctions. The sole effect of imposing sanctions for S is to reduce her payoff up to that subgame by $c(\sigma)$, while the sole effect of not imposing sanctions is to reduce her payoff up to that point by either 0 (if her demand was met) or ϕ (if it was not). If $\sigma < \overline{\sigma}$, then because $c'(\sigma) > 0$, it must be that $0 < c(\sigma) < \phi$ and subgame perfection implies that S will impose sanctions if and only if her demand was not met in the prior play.

If instead $\sigma > \overline{\sigma}$, then it must be that $c(\sigma) > \phi > 0$ and subgame perfection implies that S will not impose the sanctions in equilibrium. Backward induction implies that T sets $x = 0$ if he wins an internal political contest (IPC) while R would set $x = r$ if victorious, so that the continuation values of an IPC for the two players are $V^T_{IPC} = p(1-r) + 1 - p - \gamma_T$ and $V^R_{IPC} = p + (1-p)(1-r) - \gamma_R$. If instead T makes an offer x that R accepts, their continuation values are $V^T_x = 1 - x$ and $V^R_x = 1 - |r - x|$. Subgame perfection requires R to accept any $x > pr - \gamma_R$ and reject any $x < pr - \gamma_R$, so that T's optimal offer is $x = \max\{pr - \gamma_R, 0\}$. Since neither player's equilibrium strategy depends on d, both ignore S's demand. □

Proposition 1. *An internal political contest occurs if and only if* $d \geq p\max\{r, d\} + (1-p)\sigma + \gamma_T$, $p\sigma \geq \gamma_R + \gamma_T + 2p\max\{d - r, 0\}$, *and* $\sigma \leq \overline{\sigma}$. *If R wins, she will set policy* $x = \max\{r, d\}$ *and no sanctions will be imposed. If T wins, he will set policy* $x = 0$ *and sanctions will be imposed.*

Proof. If $\sigma > \overline{\sigma}$ or $d = 0$, then the proof of Lemma 1 establishes that no IPC will occur, so assume $\sigma \leq \overline{\sigma}$ and $d > 0$ for the remainder of the proof. There are three cases to deal with depending on how large σ is relative to $d - r$ and d. We ignore the edge cases $\sigma = d$ and $\sigma = d - r$. In those cases, two equilibria exist: one entails the behavior we demonstrate for a greater σ, the other that for a lower σ.

Case 1: $\sigma < d - r < d$. After winning an IPC, T would choose $x = 0$ and R would set $x = r$. Then the values of an IPC are $p(1-r) + (1-p)(1-\sigma) - \gamma_T$ for T and $p(1-\sigma) + (1-p)(1-r) - \gamma_R$ for R. Then it is easily shown that in the bargaining prior to an IPC, R and T would strictly prefer accepting and offering some $x > pr - p\sigma - \gamma_R$ respectively to an IPC, so that no IPC can occur in equilibrium.

Case 2: $\sigma > d$, then T would set $x = d$ and R would set $x = \max\{r, d\}$ after winning an IPC. Then the values of an IPC are $p(1 - \max\{r, d\}) + (1-p)(1-d) - \gamma_T$ for T and $p(1 - \max\{d - r, 0\}) + (1-p)(1 - |r - d|) - \gamma_R$ for R. If $d \geq r$, then clearly R and T would

strictly prefer accepting and offering $x = d$ respectively, since this saves both their respective cost of an IPC while yielding the same policy. If instead $d < r$, then it is easily shown that R and T would strictly prefer respectively accepting and offering some $x > pr + (1-p)d - \gamma_R$ to an IPC, so that no IPC can occur in equilibrium.

Case 3: $d - r < \sigma < d$, then T will set $x = 0$ and R will set $x = \max\{r, d\}$ after winning an IPC. Then the values of an IPC are $p(1 - \max\{r, d\}) + (1 - p)(1 - \sigma) - \gamma_T$ for T and $p(1 - \max\{r, d\} + r) + (1 - p)(1 - r) - \gamma_R$ for R. In bargaining prior to an IPC, the minimum x which R could possibly accept in equilibrium is $x_{min} \equiv p(2r - \max\{r, d\}) - \gamma_R$.

First suppose $d \leq x_{min}$, which implies $x_{min} \geq 0$. Observe that the only possible optimal offers for T are x_{min} and any $x < x_{min}$, if $x_{min} > 0$. Any offer $x \geq x_{min}$ will be accepted by R and not sanctioned by S; the best of these for T is clearly $x = x_{min}$. Any offer below x_{min} will lead to an IPC with the values above. Hence T chooses effectively between an IPC and offering $x_{min} = pr - \gamma_R$ (since $d \leq x_{min} \leq pr - \gamma_R < r$). The latter would yield a value of $1 - pr + \gamma_R$ for T and so would be strictly preferred. Thus no IPC can occur in equilibrium.

Now suppose instead that $d > x_{min}$. Observe that the only possible optimal offers for T are d, $\max\{0, x_{min}\}$, and any $x < x_{min}$ if $x_{min} > 0$. Any offer x that is at least d would be accepted by R and not sanctioned by S, yielding a continuation value for T of $1 - x$; of these, $x = d$ is strictly best for T. Any offer in $[x_{min}, d)$ would be accepted but sanctioned, giving a value for T of $1 - x - \sigma$; of these, $x = x_{min}$ is strictly best for T. Any offer less than x_{min} would lead to an IPC with the value above.

The first inequality in the proposition is equivalent to the condition that T would strictly prefer an IPC to yielding to S's demand, and the second inequality is equivalent to the condition that T would strictly prefer an IPC to satisfying R while spurning S's demand. Hence, in equilibrium T will spurn S's demand and induce an IPC with an unsatisfactory offer to R if and only if both inequalities hold.

This proves that the inequalities in the proposition are necessary for an IPC to occur in

equilibrium. Those inequalities are sufficient when combined with the conditions $d-r<\sigma<d$ and $d>x_{min}>0$ used in the case arguments above. We proceed to show that these latter conditions are not necessary because they are implied by the inequalities in the proposition.

From the first inequality in the proposition, we have $d\geq pd+(1-p)\sigma+\gamma_T>\sigma$, and also $d\geq pr+(1-p)\sigma+\gamma_T>pr-\gamma_R\geq x_{min}$. From the second inequality in the proposition, we have $p\sigma\geq 2p(d-r)\Rightarrow\sigma>d-r$. This leaves only $x_{min}>0$. To see this, first suppose $d\geq r$. Then the second inequality can be rearranged to $2pr\geq 2pd-p\sigma+\gamma_R+\gamma_T$. Because the first inequality implies $d>\sigma$, we have $2pr>pd+\gamma_R+\gamma_T\Rightarrow x_{min}=2pr-pd-\gamma_R>\gamma_T\geq 0$. Now suppose $d<r$. Because the first inequality implies $d>\sigma$, we have $r>\sigma$, so that the second inequality implies $pr>\gamma_R+\gamma_T$ and thus $x_{min}>0$. □

Proposition 2. *Suppose $r=1$. If $p\overline{\sigma}\geq\gamma_R+\gamma_T$ and $pk>(1-p)\left[\overline{\sigma}+c\left(\frac{\gamma_R+\gamma_T}{p}\right)\right]+\gamma_S+\gamma_T$, then S will make a demand that she knows will be spurned and threaten sanctions of severity $\sigma=\frac{\gamma_R+\gamma_T}{p}$, and an internal political contest will occur. Otherwise, S will demand $d=\max\{\overline{\sigma},\min\{p-\gamma_R+\overline{\sigma},p+(1-p)\overline{\sigma}+\gamma_T\}\}$ and threaten sanctions of severity $\overline{\sigma}$, T will yield to S's demand, and no internal political contest or sanctions will occur.*

Proof. For any σ,d S chooses, either an IPC will result or T will make an offer that R accepts. In order to determine S's best option, we compute the optimal choice for S among those that will lead to an IPC, and compare the value for S of this to that of the optimal choice that will lead to an accepted offer.

S's best choice among those that will lead to an IPC is immediate from Proposition 2: $\sigma=\frac{\gamma_R+\gamma_T}{p}$, which must be no greater than $\overline{\sigma}$ to cause an IPC, and $d\geq p+(1-p)\sigma+\gamma_T$, which yield a continuation value for S of $V^S_{IPC}=p+(1-p)\left[-c\left(\frac{\gamma_R+\gamma_T}{p}\right)-k\right]-\gamma_S$. Any σ or d lower than these will, by Proposition 1, not lead to an IPC. Any σ higher than this that still leads to an IPC will change S's value only by increasing the cost of the sanctions S will have to impose if T wins the IPC, making S strictly worse off.

We will partition the possible choices by S that do not lead to an IPC and compute the optimal choice within each partition.

Case 1: $\sigma > \overline{\sigma}$: By the proof of Lemma 1, T offers $\max\{p-\gamma_R, 0\}$ and R accepts, giving S a value $\max\{p-\gamma_R, 0\} - k$.

Case 2: $d \leq \sigma \leq \overline{\sigma}$: This implies that $d < p + (1-p)\sigma + \gamma_T$, so by Proposition 1, no IPC will occur. Observe that T would set $x = d$ after winning an IPC, since instead setting $x = 0$ would result in suffering the cost of sanctions σ at a gain of $d \leq \sigma$. R's reservation value is thus $V'^R_{IPC} = p + (1-p)d - \gamma_R$ and T's offer will be at least this. If $V'^R_{IPC} < d$, then T would strictly prefer to offer d, since offering less would result in suffering sanctions at a cost σ for a gain of at most d. Thus T will make an offer of $x^* = \max\{V'^R_{IPC}, d\}$, giving a value for S of $x^* - k$. The choice that maximizes this value is $d = \sigma = \overline{\sigma}$, yielding $\max\{p + (1-p)\overline{\sigma} - \gamma_R, \overline{\sigma}\} - k$ for S. This is strictly better for S than the previous partition.

Case 3: $\sigma \leq \overline{\sigma}$, $\sigma < d < \min\{p + \sigma - \gamma_R, p + (1-p)\sigma + \gamma_T\}$: Observe that this partition only exists if $\sigma < p + \sigma - \gamma_R \Leftrightarrow p - \gamma_R > 0$. The second term in the minimum implies that the first inequality in Proposition 1 does not hold, ruling out an IPC. Because $d > \sigma$, T would set $x = 0$ after winning an IPC so that R's reservation value is $V^R_{IPC} = p - \gamma_R$. If $d \leq p - \gamma_R$, R's minimum also satisfies S and so is clearly T's optimal offer. If $d > p - \gamma_R$, offering R's minimum would give T $1 - p + \gamma_R - \sigma$ while offering d would give T $1 - d$, which by definition of the partition is larger, so that offering d is optimal. Hence S's optimal choice is to set $\sigma = \overline{\sigma}$ and $d^* = \min\{p + \overline{\sigma} - \gamma_R, p + (1-p)\overline{\sigma} + \gamma_T\}$, giving her a value of $d^* - k$.

When this partition exists, then both terms in the minimum in S's value here exceed both terms in the maximum in S's value in the previous case, so that the best of this partition is strictly better for S than anything in the previous partitions.

Case 4: $\sigma \leq \overline{\sigma}$, $\sigma < d$, $d > \min\{p + \sigma - \gamma_R, p + (1-p)\sigma + \gamma_T\}$: The definition of this partition implies that the first two of the three inequalities in Proposition 1 hold, so for an IPC not to occur, it must be that the the last inequality in that proposition does not hold:

$\sigma < \frac{\gamma_R+\gamma_T}{p}$. This in turn implies that the first term in the minimum defining this partition is smaller, so that $d > p+\sigma-\gamma_R$. Because $\sigma < d$, T would set $x = 0$ after winning an IPC and R's reservation value is $V^R_{IPC} = p-\gamma_R$. If T offered d, he would receive a value of $1-d$, while offering R's minimum (or 0, if R's minimum is not feasible) would give him $1-p+\gamma_R-\sigma$ (or $1-\sigma$ if R's minimum is infeasible), which is larger and so strictly preferred. Hence S would receive a value of $\max\{p-\gamma_R, 0\} - c(\sigma) - k$, which is strictly worse for S than the best of any previous partition.

Collecting the analysis, if $p-\gamma_R \leq 0$, it is not possible for S to incite an IPC, and the optimal choice for S is $d = \sigma = \overline{\sigma}$, yielding a value of $\max\{p+(1-p)\overline{\sigma}-\gamma_R, \overline{\sigma}\} - k$. If instead $p-\gamma_R > 0$, then the optimal choice for S that does not incite an IPC is $\sigma = \overline{\sigma}$, $d = d^*$, giving her a value of $\min\{p+\overline{\sigma}-\gamma_R, p+(1-p)\overline{\sigma}+\gamma_T\} - k$. Comparing this value to that of the optimal IPC-inciting choice from above, and recalling that by Proposition 1 inciting an IPC is only feasible if $p\overline{\sigma} \geq \gamma_R+\gamma_T$, yields the conditions in the proposition. The requirement for an IPC to be possible that $p-\gamma_R > 0$ is implied by the first condition in the proposition, so that those conditions are sufficient. □

Proposition 3. *If R's ideal policy is close enough to S's (r close enough to 1), then S becomes more likely to use sanctions to incite regime change rather than policy change as the cost of coexisting with T (k) rises, the costs of an internal political contest (γ_R, γ_T, γ_S) fall, R's ideal policy moves closer to S's (r rises), and the probability of R winning (p) rises.*

Proof. We proceed to calculate the best possible choice for S of a d, σ among pairs that will not incite an IPC. We then examine how the feasibility of inciting an IPC varies in the parameters, and compute the best possible choice among strategies that will incite an IPC when any exist. We then compare the continuation values of the two to obtain the desired comparative statics. We partition S's possible strategies into subsets where the effect of S's strategy on subsequent behavior by T and R is clear.

Case 1: $\sigma > \overline{\sigma}$: Similarly to Proposition 2, S gets a value of $\max\{pr - \gamma_R, 0\} - k$.

Case 2: $d \leq \sigma \leq \overline{\sigma}$: Similarly to Proposition 2, S gets a value of $\max\{pr + (1-p)\overline{\sigma} - \gamma_R, \overline{\sigma}\} - k$, strictly better than the previous partition.

Case 3: $\sigma \leq \overline{\sigma}$, $\sigma < d - r$: Observe that T would set $x = 0$ and R would set $x = r$ after winning an IPC, since for both the value of this outweighs the value of meeting S's demand and avoiding sanctions. R's expected value from an IPC is thus $V^R_{IPC} = p(1-\sigma) + (1-p)(1-r) - \gamma_R$, and the minimum offer x needed to satisfy R is given by $1 - |r - x| = V^R_{IPC}$, so that $x = pr - p\sigma - \gamma_R$. It is easy to check that T strictly prefers making this offer (or 0, if this offer is infeasible) to any other, so that S gets $\max\{pr - p\sigma - \gamma_R, 0\} - c(\sigma) - k$, strictly worse than the previous partition.

Case 4: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d \leq p\left[2r - \max\{r, d\}\right] - \gamma_R$: Observe that in this and all subsequent partitions, T would set $x = 0$ after winning an IPC, while R would choose $x = \max\{r, d\}$. The minimum offer that R would accept in equilibrium is $x_{min} \equiv p\left[2r - \max\{r, d\}\right] - \gamma_R$. The last inequality defining this partition means that offering x_{min} is feasible and would also meet S's demand (since $x_{min} \geq d \geq 0$). It also implies that $d \leq pr - \gamma_R < r$, so that $x_{min} = pr - \gamma_R$. It is easy to show that T would strictly prefer making this offer to any other, yielding $pr - \gamma_R - k$ for S, strictly worse than partition 2 when the current partition exists.

Case 5: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d$, $p\left[2r - \max\left\{r, d\right\}\right] - \gamma_R < d$
$\leq \min\left\{p\left[2r - \max\left\{r, d\right\}\right] + \sigma - \gamma_R, p\max\left\{r, d\right\} + (1-p)\sigma + \gamma_T\right\}$: The second-to-last inequality implies a policy of x_{min} would not meet S's demand, so that in equilibrium there are only three possible optimal offers for T: d, which satisfies both S and R; x_{min}, which satisfies R but leads to imposed sanctions; and anything less than x_{min}, which leads to an IPC. The first term in the minimum implies that T would prefer offering d to x_{min}, while the second term implies T would prefer d to an IPC. Hence in equilibrium d is offered and accepted, so that S gets a value of $d - k$.

Throughout the remainder of the proof, we will interpret "R's ideal policy is close enough to S's" to mean that $r > \overline{\sigma} + \frac{\gamma_T}{1-p}$. This implies that $pr + (1-p)\sigma + \gamma_T \leq r$ for any $\sigma \leq \overline{\sigma}$, which in turn means the inequality that distinguishes this partition can be simplified to $pr - \gamma_R < d \leq \min\{pr + \sigma - \gamma_R, pr + (1-p)\sigma + \gamma_T\}$. Since both terms in the minimum rise in σ and are independent of d, the optimal strategy for S within this partition must be $\sigma = \overline{\sigma}$ and $d = \min\{pr + \overline{\sigma} - \gamma_R, pr + (1-p)\overline{\sigma} + \gamma_T\}$.

Case 6: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d$, $p\sigma \leq \gamma_R + \gamma_T - 2p\,[r - \max\{r, d\}]$, $d > \min\{p\,[2r - \max\{r, d\}] + \sigma - \gamma_R, p\max\{r, d\} + (1-p)\sigma + \gamma_T\}$: Note that the second-to-last inequality implies that the first term in the minimum binds. The last inequality then implies that T would strictly prefer offering x_{min} and suffering sanctions to yielding to S's demand, while the second-to-last implies that T would strictly prefer offering x_{min} (or 0, if $x_{min} < 0$) and suffering sanctions to a lesser offer that would be rejected by R. Hence in equilibrium $\max\{x_{min}, 0\}$ is offered and accepted. By the last inequality, this does not meet S's demand, so that S has to impose sanctions and receives a value of $\max\{x_{min}, 0\} - c(\sigma) - k \leq \max\{pr - \gamma_R, 0\} - c(\sigma) - k$, strictly worse than partition 2.

Case 7: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d$, $p\sigma > \gamma_R + \gamma_T - 2p\,[r - \max\{r, d\}]$, $d > \min\{p\,[2r - \max\{r, d\}] + \sigma - \gamma_R, p\max\{r, d\} + (1-p)\sigma + \gamma_T\}$: Note that the second-to-last inequality implies that the binding term in the minimum is the second one. Then the conditions of Proposition 1 are met and this partition results in an IPC, yielding a value for S of $p\max\{r, d\} + (1-p)\,[-k - c(\sigma)] - \gamma_S$.

From the arguments above, it is clear that either partition 2 or 5 contains the best non-IPC-inciting strategy for S, and that partition 7 contains the best IPC-inciting strategy for S. Observe that which of 2 or 5 contains the optimum non-IPC-inciting strategy for S is determined by whether $pr - \gamma_R$ is positive or not. If $pr - \gamma_R < 0$, then partition 5 does not exist since it requires $d \leq pr + \sigma - \gamma_R < \sigma$, but also $d > \sigma$, so it must be that the best of partition 2 is optimal overall. If $pr - \gamma_R > 0$, then the first term in the minimum

for partition 5 clearly exceeds both terms in the maximum defining partition 2. The second term in partition 5's minimum clearly exceeds the second term in partition 2's maximum, and also exceeds the first term since by assumption $r > \overline{\sigma}$. Hence the optimal non-IPC-inciting strategy for S is $\sigma^*_{non} \equiv \overline{\sigma}$ and:

$$d^*_{non} \equiv \begin{cases} \overline{\sigma} & \text{if } pr - \gamma_R \leq 0 \\ \min\left\{pr + \overline{\sigma} - \gamma_R, pr + (1-p)\overline{\sigma} + \gamma_T\right\} & \text{if } pr - \gamma_R > 0 \end{cases}$$

Now consider the feasibility for S of instead inciting an IPC. Using our assumption that $r > \overline{\sigma} + \frac{\gamma_T}{1-p}$, we have that $pr + (1-p)\sigma + \gamma_T < r$ for any $\sigma \leq \overline{\sigma}$, which implies that the first condition for an IPC to occur from Proposition 1 can always be satisfied by choosing $d = r$ and $\sigma \leq \overline{\sigma}$. With $d = r$, the second condition from Proposition 1 can be written $p\sigma > \gamma_R + \gamma_T$. Combining this with the third and final condition that $\sigma \leq \overline{\sigma}$, it is possible in equilibrium for S to incite an IPC if and only if $p\overline{\sigma} \geq \gamma_R + \gamma_T$. Clearly, the feasibility of inciting an IPC rises in p and falls in γ_R and γ_T, and does not depend on k, γ_S, or r, subject to our assumption that r is high enough.

Next we compute the optimal IPC-inciting strategy for S, given that such a strategy is feasible. From the above argument, if any strategy will incite an IPC, then setting $d = r$ and $\sigma = \frac{\gamma_R + \gamma_T}{p}$ will do so, so the question is simply whether S would do better to demand more than r when trying to incite an IPC. Given that $d^*_{IPC} \geq r$, the first condition for an IPC to occur from Proposition 1 is automatically satisfied since $d > pd + (1-p)\sigma + \gamma_T \Leftrightarrow d > \sigma + \frac{\gamma_T}{1-p}$, which by assumption is less than r. From partition 7, the continuation value for S will be $pd + (1-p)\left[-k - c(\sigma)\right] - \gamma_S$. Since this value rises in d but falls in σ, the optimal strategy must entail the second feasibility condition being just met, or $p\sigma = \gamma_R + \gamma_T + 2p(d - r)$. Substituting this for σ in S's continuation value and differentiating with respect to d, we obtain the first-order condition $c'\left(\frac{\gamma_R + \gamma_T}{p} + 2(d^* - r)\right) = \frac{p}{2(1-p)}$.

The third and final feasibility condition for an IPC to occur from Proposition 1 requires $\sigma \leq \overline{\sigma}$. The argument inside $c'(\cdot)$ above will be equal to this upper bound on σ when $d^* = r + \frac{\overline{\sigma}}{2} - \frac{\gamma_R+\gamma_T}{2p}$, so that this is an upper bound on the optimal demand. Using this and the first-order condition above, we have $\sigma^*_{IPC} = \frac{\gamma_R+\gamma_T}{p} + 2(d^*_{IPC} - r)$ and:

$$d^*_{IPC} = \begin{cases} r & \text{if } \frac{p}{2(1-p)} \leq c'\left(\frac{\gamma_R+\gamma_T}{p}\right) \\ c'\left(\frac{\gamma_R+\gamma_T}{p} + 2(d^*_{IPC} - r)\right) = \frac{p}{2(1-p)} & \text{if } c'\left(\frac{\gamma_R+\gamma_T}{p}\right) < \frac{p}{2(1-p)} \leq c'\left(\overline{\sigma}\right) \\ r + \frac{\overline{\sigma}}{2} - \frac{\gamma_R+\gamma_T}{2p} & \text{if } c'\left(\overline{\sigma}\right) < \frac{p}{2(1-p)} \end{cases}$$

Finally, we can compare the continuation values of the best non-IPC-inciting and best IPC-inciting strategies for S, given that the latter is feasible. We showed above that feasibility entails $p\overline{\sigma} \geq \gamma_R + \gamma_T$. Since $r > \overline{\sigma}$ by assumption, this inequality also implies that $pr > \gamma_R$. When inciting an IPC is feasible, we can use these to simplify the optimal non-IPC-inciting strategy and write its continuation value as:

$$V^S_{non} = d^*_{non} - k = pr + (1-p)\overline{\sigma} + \gamma_T - k$$
$$V^S_{IPC} = pd^*_{IPC} + (1-p)\left[-k - c\left(\sigma^*_{IPC}\right)\right] - \gamma_S$$

Clearly, inciting an IPC becomes more attractive to S as k rises and γ_S falls. An increase in either γ_R or γ_T lowers d^*_{IPC} or leaves it unchanged in all the cases in its definition, and shifts the constraints defining those cases in a way that also decreases d^*_{IPC} or leaves it unchanged, and raises σ^*_{IPC} even holding d^*_{IPC} constant, so that V^S_{IPC} must decline. By contrast, V^S_{non} clearly rises in γ_T and does not depend on γ_R, so that inciting an IPC becomes more attractive to S as γ_R or γ_T fall.

Next consider an increase in r. We can see that the derivative of d^*_{IPC} with respect to r is 1, while that of σ^*_{IPC} is 0 accounting for its dependence on d^*_{IPC}. Hence V^S_{IPC} rises in r

at a rate p, just like V^S_{non}. However, recall that k is treated as a function of r, and assumed to be increasing in r in the model setup. Hence, as r grows, inciting an IPC becomes more attractive to S.

Lastly consider an increase in p. Differentiating V^S_{IPC} with respect to p, we have:

$$\begin{aligned}\frac{\partial V^S_{IPC}}{\partial p} &= d^*_{IPC} + p\frac{\partial d^*_{IPC}}{\partial p} + k + c\left(\sigma^*_{IPC}\right) - (1-p)c'\left(\sigma^*_{IPC}\right)\frac{\partial \sigma^*_{IPC}}{\partial p} \\ &= d^*_{IPC} + p\frac{\partial d^*_{IPC}}{\partial p} + k + c\left(\sigma^*_{IPC}\right) - (1-p)c'\left(\sigma^*_{IPC}\right)\left[-\frac{\gamma_R+\gamma_T}{p^2} + 2\frac{\partial d^*_{IPC}}{\partial p}\right]\end{aligned}$$

Observe that d^*_{IPC} cannot fall: in every case in its definition, it increases or is unchanged, and the constraints shift in favor of higher cases. At $d^*_{IPC} = r$, $\frac{\partial d^*_{IPC}}{\partial p} = 0$, and every term in the above expression is non-negative, so that $\frac{\partial V^S_{IPC}}{\partial p} > d^*_{IPC} = r > r - \overline{\sigma} = \frac{\partial V^S_{non}}{\partial p}$. If instead $d^*_{IPC} > r$, then from its definition we must have $c'\left(\sigma^*_{IPC}\right) \leq \frac{p}{2(1-p)}$, implying:

$$\begin{aligned}\frac{\partial V^S_{IPC}}{\partial p} &\geq d^*_{IPC} + p\frac{\partial d^*_{IPC}}{\partial p} + k + c\left(\sigma^*_{IPC}\right) - (1-p)\frac{p}{2(1-p)}\left[-\frac{\gamma_R+\gamma_T}{p^2} + 2\frac{\partial d^*_{IPC}}{\partial p}\right] \\ &= d^*_{IPC} + k + c\left(\sigma^*_{IPC}\right) + \frac{\gamma_R+\gamma_T}{2p} > r - \overline{\sigma} = \frac{\partial V^S_{non}}{\partial p}\end{aligned}$$

Hence inciting an IPC becomes more attractive to S as p rises.

Since the attractiveness of inciting an IPC rises in p, r, and k and falls in γ_S, γ_R, and γ_T, and its feasibility rises in in p, falls in γ_R and γ_T, and does not depend on the other variables, we have established the claims in the proposition. □

2 Chile: Overthrowing a Democracy

The US imposed sanctions on Chile following the 1970 election victory of Salvador Allende, at the head of a coalition of socialist and communist parties. Three years later, the Allende regime was overthrown and replaced by a military junta and the US dropped its sanctions.

Primary sources show that the US decision calculus and anticipated reactions within Chile adhered quite closely to the predictions of our model.[1]

The US government perceived Allende to have interests sharply conflicting with its own, but the Chilean opposition to be much more favorable to the US. Just before the election, the US anticipated that "There would be problems for US-Chile relations under either [of Allende's opponents] [...] If Allende wins, the problems created for the US would be much greater" (C47). The US also assessed that the opposition would not contest Allende's power absent US intervention. "There is no reason to believe that the Chilean armed forces will unleash a civil war or that any other intervening miracle will undo his victory" (C62). Instead, Allende was expected to set policies that compromised just enough with the opposition to avoid a contest, just as the model predicts. Allende's government "would seek to destroy, neutralize, or obtain the support of various groups and institutions" in the opposition by changing policies slowly and undermining their political strength, and "a sufficiently gradualist and skillful approach by Allende could avoid provoking the military almost indefinitely" (CS13). Nonetheless, change would be "as rapid as possible without inciting a dangerous reaction from the military" (CS33). Finally, the US believed that only serious degradation of Chile's economy could stop this accommodation and lead to an internal political contest: "economic conditions might provide the essential provocation [...] that might set in motion the politico-military forces to defeat Allende" (C68).

Facing a disfavored regime and a favored opposition that would have to be incited to contest power, the US quickly arrived at the two possible best strategies our model predicts. As described in a series of interagency strategy papers (CS13, CS14, CS30, C166, CS33), it could bargain with Allende, influencing his policies with the threat of sanctions but not attempting to incite his overthrow ("Option B" in CS13). Alternatively, it could ap-

[1]All primary sources are drawn from McElveen and Siekmeier (2014) (with sources therein cited as "C[document number]" for brevity) and its supplement, McElveen and Siekmeier (2015) (with sources therein cited as "CS[document number]").

ply sanctions regardless of Allende's policies, possibly supplemented with covert support to the opposition, to prevent him from consolidating power and possibly lead to his overthrow ("Option C," with covert support described in CS14).[2]

The debate between these two strategies turned on the the feasibility of inciting a coup (the first condition of Proposition 2) and the probability it would succeed (p), the costs for the US of an incited coup (γ_S), and the costs of living with Allende (k), in line with Proposition 3. Those officials who preferred coercive sanctions worried that "there is little chance that, even with our stimulation [by covert support], an overthrow of Allende would be attempted, and there is almost no way to evaluate the likelihood that such an attempt would be successful even were it made" and recognized that "an unsuccessful attempt [...] would have grave consequences for our relations in Chile, in the hemisphere, in the United States and elsewhere in the world" (CS14). These officials "view[ed] the risks that Allende [would] consolidate himself and the long-range consequences therefrom [as] less dangerous to us than the immediate probable reaction to attempts to oppose Allende" (C172). By contrast, officials who advocated incitement sanctions believed the US "[could] affect events, and that the risks of stirring up criticism to our position elsewhere are less dangerous to us than the long-term consolidation of a Marxist government in Chile" (C172). This view was strengthened by the realization that US incitement could be undertaken quietly so as to minimize its cost to the US and the danger that it would only lead to entrenchment of the Allende regime (C172).

The President chose incitement, ordering US officials to "make the economy scream" (C93) prior to Allende's inauguration.[3] The US Ambassador told Chile's then-president that "not a nut or bolt will be allowed to reach a Chile under Allende [...] Once Allende

[2] "Option A," entailing simply living with Allende without coercion, was quickly discarded (C150).

[3] The US also provided covert support to the opposition in an effort to maintain its viability. However, the US assessed that its "support would be a helpful but marginal factor in the calculations of the Chilean military in an attempt to topple an Allende administration" (CS14).

comes to power, we shall do all within our power to condemn Chile and the Chileans to utmost deprivation and poverty, a policy designed for a long time to come to accelerate the harsh features of a Communist society in Chile" (C108). The US understood that Allende would not concede enough to avoid these sanctions. "It is clear that Allende is not voluntarily going to modify his goals, nor is he likely to have any interest in negotiating such a modification just to get along with us" (C159).

As the model predicts, US sanctions (combined with the economic consequences of Allende's own policies) caused increasingly serious economic difficulties in Chile, which in turn undermined the stability of the Allende regime (C270, 288, 294, 302, 307, 341). The US Ambassador observed that "Politically, Chile seems to be experiencing increasing confrontation and polarization of political forces. Sharpening economic crisis has stimulated a stronger opposition" (C276). A national intelligence estimate agreed that "The decline in Allende's political fortunes has in good part been caused by the worsening economic siutation" (CS116).

Three attempts at a coup occurred (C162, 168, 334, 346, 347), of which the last succeeded. Afterwards, the national security advisor said to the President "We didn't do it. I mean we helped them. [Redacted] created the conditions as great as possible" (C357). As the model predicts, with a more favorable regime now in power, the US moved immediately to express support and end its sanctions (C349, 353, 355, 361).

3 Iraq: Replacing a Dictator

The US and UN imposed comprehensive sanctions on Iraq following its 1990 invasion of Kuwait and maintained them after its forces were expelled. Internal attempts at regime change occurred but failed, and Saddam Hussein remained in power until the US invaded Iraq and deposed him in 2003, after which sanctions were dropped. We show that the US

decision calculus and subsequent behavior accords with our model.[4]

As with South Africa, the US debated the costs of coexisting with Saddam's Iraq (k) in deciding whether to use coercive or incitement sanctions. "On the one hand, there were those who believed that with Iran defeated there was no need to look the other way at Saddam's human rights abuses, his continued [...] support for international terrorist organizations, and his ravenous appetite for weapons of mass destruction. [...] On the other hand, numerous voices [...] had been arguing [...] that although Saddam was not perfect, he was pragmatic and broadly shared U.S. goals in the region, and so Washington should pursue a policy of engagement to wean him from some of his more egregious sins and allow him to take the vacant seat of the shah as America's proxy in the Persian Gulf" (P838). The Bush administration chose coercive sanctions: "carrots (political and economic incentives) would be used to entice the Iraqis into ending their pursuit of WMD and human rights abuses, while holding out the stick of sanctions if they did not mend their ways" (P855).

Iraq's invasion of Kuwait increased estimates of k and led to incitement sanctions. "The Iraqi invasion was a nasty shock for the Bush administration. It represented a serious threat to America's principal objectives in the Persian Gulf region" and "demonstrated that the [...] policy of constructive engagement [i.e., coercive sanctions] and its assessment that Saddam was 'pragmatic' and 'moderate' were mistaken" (P1018). "Saddam had now been revealed as an extremely dangerous leader, and the administration recognized that the past revelations regarding Iraq's unshakable pursuit of weapons of mass destruction and outrageous violations of human rights were further proof that the Baghdad regime was a force for real instability in the vital Persian Gulf region" (P1057). "The Bush administration [...] persuaded the U.N. Security Council to pass a series of resolutions [...] demanding that Iraq withdraw, and imposing severe sanctions on Iraq for failing to comply" (P1066), but

[4]We rely heavily on the account in Pollack (2002), written by a key analyst and later official of the first Bush and Clinton administrations (cited as P[page number]), as well as the intelligence estimates contained in Richelson (2005) (cited as R[document number]).

"no one [in the US government] believed the sanctions would have been adequate to force Saddam to give up Kuwait by themselves" (P251). Instead, "intelligence experts expected dissatisfaction with Saddam Hussein's rule to grow [...] as a global trade embargo began to crimp that nation's economy seriously" and "the Bush administration [was] pressing the worldwide embargo vigorously in part because officials believe[d] it to be the most effective means of unseating the Iraqi President" (Wines, 1990). The Bush administration later made this explicit: "Deputy National Security Advisor Robert Gates announced that 'Saddam is discredited and cannot be redeemed. His leadership will never be accepted by the world community. Therefore, Iraqis will pay the price while he remains in power. All possible sanctions will be maintained until he is gone ... Any easing of sanctions will be considered only when there is a new government.' " (P1445).

The Bush administration also evaluated the chances of success (p), policy preferences of a new government (r), and cost to the US (γ_S) of attempts to overthrow Saddam by various opposition elements in Iraq. Just prior to the expulsion of Iraq's forces from Kuwait, the US assessed that Saddam "faces formidable domestic challenges. He is certain to encounter deep public anger over the devastating economic and political effects of the crisis and resentment within the military" (R6). After, it judged that "Saddam faces his most serious political challenge in more than twenty years of power" and "could be in mortal danger" (R8). Intelligence "experts believed that [high officials] would 'present the greatest threat' to President Hussein's continued power, 'more so than those who wish they were in a greater position of power,' like dissident minority groups" (Wines, 1990). They also viewed the former as having preferences more amenable to the US and their attempt to win power as less costly. In a scenario where "Saddam is Removed By Insiders", it would be by a coup that "would not necessarily involve foreign backers" (R8). A "successor government probably would make significant concessions to reduce regional tensions in order to consolidate its own position" and the "new leaders would look to the United States for assurances of tacit

acceptance" including "that all sanctions be lifted", so that the US could influence this new government (R8). By contrast, a successful insurrection by minority groups could lead either to "a pro-Iranian Shia state" possibly featuring "an Iranian-style Islamic republic [...that] would revive Islamic revolutionary fervor in Iran and cause increasing threats of Shia subversion and terrorism to" US allies in the region, or to "a Lebanon-style power vacuum" in which "a general civil war [...] might develop" and "Iran, Syria, and Turkey [...] could be tempted to intervene" (R8).

Incitement sanctions were re-evaluated by the Bush administration after the Gulf War and by the Clinton administration when it took office, but remained attractive for the same reasons (R9–12). For example, a National Intelligence Estimate from Clinton's first year in office corresponds remarkably closely with our theory (R12). It assumed that "Saddam Husayn will not alter his basic domestic and foreign policy goals: to maintain his hold on power by any means necessary, to reimpose full control over the country [including via drastic abuses of human rights], to rebuild Iraq's military might—including weapons of mass destruction programs—and to make Iraq the dominant regional power" at the expense of the US; that is, that k would remain high. It also assumed that "Saddam Husayn will not comply with UN resolutions", that is, would not grant the sender's demands. It then evaluated sanctions according to "Key Questions: What are the prospects for the survival of Saddam Husayn's regime for another year? For three more years? [That is, p.] What role do sanctions, and the attendant economic hardship and diplomatic isolation, play in determining Saddam's survival? [Corresponding to incitement feasibility in our model.] If regime change occurs, what will be the most likely means: assassination, coup, popular uprising, opposition overthrow, or other? [Determining the cost of the internal political contest, γ_S.] What would be the characteristics and policies of likely successors [i.e., r]?"

Its key judgments directly connected sanctions to Saddam's potential overthrow: "Although sanctions by themselves will not directly topple Saddam, they have helped establish

an environment that threatens him" because by "debilitating Iraq's economy", they are "helping to erode Saddam's ability to preserve his power base by distributing favors to supportive organizations and persons" and "keep[ing] popular discontent high." "If enforcement of the sanctions continues unabated, there is a better-than-even chance that Saddam will be ousted during the next three years". By contrast, "a lifting of all sanctions probably would enable him to hold on to power indefinitely." Thus, incitement was feasible and overthrow likely, but only if the sanctions were maintained. "The Most Plausible Causes" of regime change were "Assassination [...] or a military coup" rather than a "popular uprising" with its attendant risks for the US of civil war and foreign intervention, so that γ_S was low. "If Saddam is overthrown, those responsible are likely to [...] hold positions in the military or security services" rather than be members of restive ethnic groups that might be pro-Iran or pro-secession, and "The new regime is unlikely to be quite as brutal, lawless and repressive", so that r is high enough to substantially reduce k. According with our theory, it anticipated that "The first priority of any post-Saddam regime would be to induce the United Nations to lift sanctions". Unsurprisingly given this assessment, the Clinton administration continued the incitement sanctions.

As our theory predicts, an internal political contest began, but the US supported only the opposition elements it judged more likely to result in a favorable new regime at lower cost. Observing the Shia and Kurdish rebellions that arose immediately after the Gulf War, "the administration was terrified that Iraq would fall into chaos or fragment along ethnic and religious lines, leaving a power vacuum in the region and no state to balance Iran. In Secretary of State James Baker's words, the United States did not assist the rebels [...because] 'The Shia were quite naturally perceived as being aligned with Iran, and the Kurds [...] had demanded an independent state of Kurdistan for decades' [...] since they were convinced that Saddam would be overthrown in a nice, clean military coup, they saw no need to aid the rebels" (P1268–1274). Instead the US covertly supported only a group

of former Iraqi officials who could instigate a coup and a group of Iraqi exiles that might be able to organize a popular revolt (P1542, 5954). The ethnic group rebellions of March 1991 were soon followed by an insurgency among the Marsh (Shia) Arabs; coup attempts or serious plots in May 1991, June 1992, March 1995, May 1995, June 1996, and January 1999; revolts of Sunni tribes essential to Saddam's power in 1993 and 1995; assassination attempts in December 1993, January 1994, December 1996; a terrorist campaign of bombings in 1994; and violent disputes within Saddam's own family that led to a possible coup attempt in August 1995 (P1456, 1645, 1726, 1788, 1789, 1791, 1808, 1892).

However, all of these attempts failed, and Saddam won the internal political contest, entrenching his rule. The ethnic group rebellions were crushed by Saddam's remaining loyal forces (P1295), the marshes were subjected to a counter-insurgency campaign and eventually drained to eliminate the sanctuary they offered to insurgents (R11, P2729), and "Saddam's security forces [...] foiled several major coup plots within weeks of their target date" (R12) and arrested all the conspirators (P1751, 1892). By virtue of the June 1996 "defeat of the most promising coup attempt the CIA would ever mount" (P1892), Saddam "demonstrated his own strength, crippled the Kurds, and drove the CIA-backed opposition out of the country [...] and as a result he was able to solidify his domestic political position" (P1967). Saddam "purg[ed] the [ruling] party of suspect elements and replac[ed] regional leaders" (R9), "reorganized the military and security services to provide additional protection for him [...] and his regime [...] and round[ed] up suspected coup plotters and disaffected military personnel" (R10), and conducted "brutal purges [...] of security officers whose loyalty he questioned" and "overhauled his government three times, and reshuffled cabinet ministers on at least four other occasions" (R12). By 2002, Pollack observed that "the sanctions have, in many ways, strengthened Saddam's iron grip [...] So many have tried to overthrow him and failed—and suffered execution or banishment for their troubles—that there are virtually no leaders of any stature inside Iraq who might rally Iraqis against

Saddam" (P2458), and "the external Iraqi opposition was a mess" (P2223). Pollack worried that a 2002 intelligence estimate that "a new covert action program to try to topple Saddam Hussein would have only about a 10 to 20 percent chance of succeeding"—a far cry from the better-than-even chance the CIA estimated in 1993—"might actually be too optimistic" (P5999). Nonetheless, as the theory predicts, the US maintained the sanctions, and did not end them until it had deposed Saddam by force in 2003.[5]

References

Coe, Andrew J. 2018. "Containing Rogues: A Theory of Asymmetric Arming." *The Journal of Politics* 80(4):1197–1210.

McElveen, James and James Siekmeier, eds. 2014. *Foreign Relations of the United States, 1969–1976, Volume XXI, Chile, 1969–1973*. Washington, DC: United States Government Printing Office.

McElveen, James and James Siekmeier, eds. 2015. *Foreign Relations of the United States, 1969–76, Volume E-16, Documents on Chile, 1969–1973*. Washington: Department of State.

Pollack, Kenneth M. 2002. *The Threatening Storm: The Case for Invading Iraq*. New York: Random House.

Richelson, Jeffrey. 2005. "Saddam's Iron Grip: Intelligence Reports on Saddam Hussein's Reign." National Security Archive Electronic Briefing Book No. 167.

Wines, Michael. 1990. "Embargo May Lead To Hussein's Ouster, C.I.A. Director Says." *New York Times* September 19(A11).

[5] Coe (2018) argues that the resort to war was caused by the US conclusion that the costs of coexisting with Saddam were still too high to bear.