Online Appendix: Supporting Information for *The Shadow of Deterrence: Why capable actors engage in contests short of war*

Author names redacted

2021-05-21

Contents

This appendix accompanies the paper "The Shadow of Deterrence: Why capable actors engage in contests short of war". It provides supplemental information concerning proofs for the formal model, the data set of Russian gray zone campaigns introduced in the paper, and robustness checks and alternate specifications for the statistical model.

1 Formal Model

1.1 Formal statement of assumptions

We formally express the assumption that the kinks in the P function are never activated in equilibrium. Letting $\tilde{g_C}$ and $\tilde{g_D}$ denote the optimal levels selected by C and D conditional on the actors selecting into

gray zone conflict (these are defined below), when Assumption 1 holds, the "min-max" statements in the P function will never be relevant to analysis.

Assumption 1*: In equilibrium,* $\rho_0 < P(\tilde{g_C}, \tilde{g_D}) < \rho_W$.

Based on the optimal $\tilde{g_C}$ and $\tilde{g_D}$ (solved below), this condition amounts to $\frac{\theta}{2\beta_C} - \frac{1}{2\beta_D} > 0$ and $0 < \rho_W - \rho_0 - \frac{\theta}{2\beta_C} + \frac{1}{2\beta_D}$ if $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$, and $\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D} > 0$ and $\kappa_D - \frac{1}{4\beta_D} < 0$ if $\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D} \leq \frac{\theta}{2\beta_C}$.

1.2 Proving Proposition 1

1.2.1 Equilibrium Intuition

Outside of gray zone conflict, C will prefer the status quo to initially going to war when

$$\theta\rho_0 \geq \theta\rho_W - \kappa_C$$

or

$$\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}.$$

Now we discuss the intuition of the equilibrium in the paper. Assume that C is optimally selecting a g_C^* such that the game ends in gray zone conflict (in other words assume that $w_C^* = 0$ and $g_C^* \geq 0$). Also assume that D selects an optimal g_D^* such that $g_D^* \leq g_C^*$ (this will be borne out by Assumption 1). D selects g_D^* characterized by

$$g_D^* \in argmax_{g_D \geq 0} \left\{1 - \rho_0 - g_C + g_D - \beta_D g_D^2\right\}.$$

We take first-order conditions with respect to g_D and solve the expression above to identify the optimal level of D's gray zone response g_D^*. This unique value is

$$g_D^* = \frac{1}{2\beta_D}.$$

Using the expression for g_D^*, D's utility in terms of the selected g_C^* is $U_D = 1 - \rho_0 - g_C^* + \frac{1}{4\beta_D}$.

We can then begin considering C's utility. There are two matters to consider. First, it could be that C will select an optimal g_C^* that is constrained by D's willingness to go to war. Essentially, if $g_C > \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$, then D's utility from war is greater than D's utility from gray zone conflict; thus, if C wants to remain in gray zone conflict and will be constrained by D's deterrent threat, C will select $\hat{g_C}$, where $\hat{g_C}$ is the greatest g_C that would make D indifferent between gray zone conflict and war, or

$$\hat{g_C} = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}.$$

Second, C may select an optimal g_C^* that is constrained by their own internal costs. When this is the case, C will select $\check{g_C}$, defined by the optimization

$$\check{g_C} \in argmax_{g_C \geq 0} \left\{\theta\left(\rho_0 + g_C - \frac{1}{2\beta_D}\right) - \beta_C g_C^2\right\},$$

which yields

$$\check{g_C} = \frac{\theta}{2\beta_C}.$$

Before discussing the true behavior, we highlight two things that do not happen. First, note that C will never select an g_C that provokes D to go to war in the final stage, because this is strictly worse than initially going

to war. Second, note that C will never select into gray zone conflict (i.e. set $w_R = 0$ and $g_C^* > 0$) if g_D^* as defined above is greater than g_C^* because C could do strictly better not paying the costs of war and selecting into the status quo ($g_C^* = 0$).

With this is place, if C optimally selects into gray zone conflict, C will select $g_C^* = \tilde{g_C}$, where

$$\tilde{g_C} = min\left\{\hat{g_C}, \check{g_C}\right\}.$$

We have now characterized what happens within gray zone conflict. We now need to describe how the game optimally plays out across the possibility of selecting into the status quo, war (at the onset; $w_A = 1$), or gray zone conflict. Because C moves first, this is ultimately C's choice. We can calculate C's decision within the two cases of gray zone conflict.

First, we consider the case when $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. This condition implies that the selected gray zone conflict will be constrained by D's deterrent threat and not C's internal costs. So, if C selects into gray zone conflict, C will select $g_C^* = \hat{g_C} = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. We can then express C's behavior in terms of θ. C prefers the status quo to gray zone conflict when

$$\theta\rho_0 \geq \theta\left(\rho_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2$$

or

$$\theta \leq \frac{\beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D}\right)}.$$

Note that the above derivation relies on $\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D} > 0$, lest the inequality sign would flip. This holds by Assumption 1.

Next, C prefers war to gray zone conflict when

$$\theta\rho_W - \kappa_C > \theta\left(\rho_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2$$

or

$$\theta > \frac{\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D}.$$

Note that the above derivation relies on $\frac{1}{4\beta_D} - \kappa_D > 0$, lest the inequality sign would flip. This holds by Assumption 1.

Next, we assume $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. This condition implies that the selected gray zone conflict will be constrained by C's internal costs and not D's deterrent threat. So, if C selects into gray zone conflict, C will select $g_C^* = \check{g_C} = \frac{\theta}{2\beta_C}$. We can then express C's behavior in terms of θ. C prefers the status quo to gray zone conflict when

$$\theta\rho_0 \geq \theta\rho_0 + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}$$

or

$$0 \geq \theta\left(\frac{\theta}{4\beta_C} - \frac{1}{2\beta_D}\right).$$

Next, C prefers war to gray zone conflict when

$$\theta\rho_W - \kappa_C > \theta\rho_0 + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}$$

or

$$\theta > \frac{\kappa_C}{\rho_W - \rho_0 - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}.$$

Note that the above derivation relies on $\rho_W - \rho_0 - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D} > 0$, lest the inequality sign would flip. This holds by Assumption 1.

When $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$, it is straightforward to see that, for a great enough θ, C's will declare war. We now demonstrate this for $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. We calculate

$$\frac{d}{d\theta}\left(U_C(war) - U_C(grayzone)\right) = \frac{d}{d\theta}\left(\theta\rho_W - \kappa_C - \left(\theta\rho_0 + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}\right)\right) = \quad \rho_W - \rho_0 - \frac{\theta}{2\beta_C} + \frac{1}{2\beta_D}$$

By Assumption 1, the right hand side is positive. Therefore, as θ increases, the war payoffs are rising faster than the gray zone payoffs, and for a great enough θ C will prefer war to gray zone conflict.

With all of this defined, we can characterize C's strategy in terms of θ; as θ increases, C prefers more degrees of conflict (i.e. larger g_C^*'s or war) to get what they want.

1.2.2 Equilibrium Behavior

Proposition 1A and the text below contains a more complete discussion of the equilibrium behavior characterized in Proposition 1.

Proposition 1A: *In equilibrium, the game will play out in the following manner.*

Case 1, $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$*:*

- *1.A. If* $\theta \leq \frac{\beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D}\right)}$ *and* $\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}$*, then C accepts the status quo. C selects* $w_C^* = 0$ *and* $g_C^* = 0$*, and D selects* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \rho_0$ *and* $U_C = \theta\rho_0$*.*
- *1.B. If* $\theta > \frac{\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D}$ *and* $\theta > \frac{\kappa_C}{\rho_W - \rho_0}$*, then C declares war. C selects* $w_C^* = 1$*, and payoffs are* $U_D = 1 - \rho_W - \kappa_D$ *and* $U_C = \theta\rho_W - \kappa_A$*.*
- *1.C. Otherwise, the game end in gray zone conflict where* C's *limited challenge is constrained by D's deterrent threat. C selects* $w_C^* = 0$ *and* $g_C^* = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$*, and D selects* $w_D^* = 0$ *and* $g_D^* = \frac{1}{2\beta_D}$*. Payoffs are* $U_D = 1 - \rho_W - \kappa_D$ *and* $U_C = \theta\left(\rho_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2$*.*

Case 2, $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$*:*

- *2.A. If* $\theta \leq \frac{2\beta_C}{\beta_D}$ *and* $\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}$*, then C accepts the status quo. C selects* $w_C^* = 0$ *and* $g_C^* = 0$*, and D selects* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \rho_0$ *and* $U_C = \theta\rho_0$*.*
- *2.B. If* $\theta > \frac{\kappa_C}{\rho_W - \rho_0 - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}$ *and* $\theta > \frac{\kappa_C}{\rho_W - \rho_0}$*, then C declares war. C sets* $w_C^* = 1$*. Payoffs are* $U_D = 1 - \rho_W - \kappa_D$ *and* $U_C = \theta\rho_W - \kappa_A$*.*
- *2.C. Otherwise, the game will end in gray zone conflict where* C's *limited challenge is constrained by* C's *internal efficiency. C selects* $w_C^* = 0$ *and* $g_C^* = \frac{\theta}{2\beta_C}$*, and D selects* $w_D^* = 0$ *and* $g_D^* = \frac{1}{2\beta_D}$*. Payoffs are* $U_D = 1 - \rho_0 - \frac{\theta}{2\beta_C} + \frac{1}{4\beta_D}$*, and* $U_C = \theta\rho_0 + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}$*.*

Working backwards, D will declare war for all $g_C > \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. If $g_C \leq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$, D will select $g_D = min\left\{\frac{1}{2\beta_D}, g_C\right\}$. When $g_D = \frac{1}{2\beta_D}$, D is selecting their optimal level of gray zone response based on their internal optimization. When $g_D = g_C$, it implies that D would be willing to select a greater gray zone response, but does not need to, essentially driving the political impact of C's limited challenges back to zero (at cost).

1.3 Observation 1 Discussion

Assume for now the parameters are such that the Case 1.C. conditions hold, and consider what happens when κ_D decreases. Because here C selects the greatest level of limited challenges that will not provoke D to war, C's selected g_C^* is a decreasing function of κ_D; therefore, because g_D^* is fixed, the final extent of gray zone conflict will be less. Of course, the analysis does not stop there. Improvements in D's willingness to go to war constrain how useful gray zone conflict is to R, and, within Case 1.C., C's utility is decreasing in $-\kappa_D$.[1] Thus, if κ_D becomes small enough, C will leave gray zone conflict and instead select into either accepting the status quo (entering into Case 1A) or going to war (entering into Case 1B). Additionally, it is worthwhile noting that as κ_D decreases, the condition that selects into Case 1 (over Case 2) has more slack, implying that improvements in D's willingness to go to war will keep D within Case 1.

Now assume the parameters are such that the Case 2.C. conditions hold, and consider what happens when κ_D decreases. Note that this will not change the selected g_C^* here, but it could break the inequality $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$ that determines whether the equilibrium is defined in Case 1 or Case 2. Thus, for a small enough κ_D, the conditions for Case 2 will break and the conditions for Case 1 will hold. When this happens, either the selected g_C^* is increasing in κ_D (Case 1.C.) or gray zone conflict is not selected (Case 1.A. or 1.B.).

1.4 Extension 1: Endogenous β_D

In the model in the paper, we treated D's gray zone efficiency β_D as exogenous. In some special cases or under some conditions, this may be too strong an assumption. In this section, we characterize an equilibrium for the game when D can have complete flexibility in selecting some $\beta_D \geq \underline{\beta_D} > 0$, where β_D cannot equal zero because D's costs from their gray zone response will then be undefined.[2] The key take away from this extension is that if β_D is endogenous (and its selection cost-less), then D's selection of β_D^* will be arbitrated by two properties. As the first property, it matters whether C prefers war to the status quo (formally, if C is type $\theta > \frac{\kappa_D}{\rho_W - \rho_0}$), or C prefers the status quo to war ($\theta \leq \frac{\kappa_D}{\rho_W - \rho_0}$). When C prefers the status quo to war, then D is in a position where D can, by selecting a low enough β_D, influence C to stop undertaking limited challenges and select into the status quo. Intuitively, when D is very good at gray zone conflict, D would select a high g_D^*, which makes gray zone conflict less productive for C. But, when C prefers war to the status quo, then D could pressure C to stop undertaking limited challenges, but this will result in C going to war with D.

As the second property, D's decision will also be arbitrated by whether D can select a gray zone efficiency β_D^* that pushes C into a level of gray zone conflict where the deterrent threat does not bind. Recall that if C optimally conducts gray zone conflict, C selects $g_C^* = min\{\hat{g_C}, \check{g_C}\}$, implying that C will either select an optimal $g_C^* = \check{g_C} = \frac{\theta}{2\beta_C}$ based on their own internal cost-benefit analysis, or select an optimal $g_C^* = \hat{g_C} = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$ tailored to make D indifferent between war and gray zone conflict (where the deterrent threat binds), with C ultimately choosing the smaller of the two. This means that if D can select a small enough β_D so that $\check{g_C} < \hat{g_C}$, then C will selecting a level of limited challenge that is below the point that would make D indifferent between war and gray zone conflict, thus granting D some surplus.

The above two properties interact. D will always prefer the status quo to gray zone conflict where the deterrent threat doesn't bind, and gray zone conflict where the deterrent threat doesn't bind to gray zone conflict where the deterrent threat does bind or war. Proposition A identifies how D selects β_D^* in one possible equilibrium. Note that this is not the only possible equilibrium.[3]

Proposition A. *As one equilibrium, in the game with endogenous β_D, D will select the following levels of β_D^*:*

[1] This follows from $\frac{d}{d\kappa_D} U_D = \theta - 2\beta_C \left[\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right] > 0$, as determined by the conditions for Case 1 to hold.

[2] For ease, we will assume that all parameters imply that the selected equilibrium is such that the selected β_D^* is strictly greater than $\underline{\beta_D}$.

[3] Consider the equilibrium space for the range of θ where the selected β_D will either push C into war or gray zone conflict where the deterrent threat binds. In the figure below, this is the far right region of the graph. Here D can select any β_D and it will grant D the same final expected utility of their wartime utility.

Case 1: $\theta \leq \frac{\kappa_D}{\rho_W - \rho_0}$*:*

- *1.A. We define* $\tilde{\beta_D}$ *as* $\theta = \frac{2\beta_C}{\tilde{\beta_D}}$ *. So long that* $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\tilde{\beta_D}}$*, then D selects* $\beta_D^* = \tilde{\beta_D}$*. The game will proceed as defined in Proposition 1, Case 2.A., where the final outcome is the status quo.*
- *1.B. Otherwise, D selects* $\beta_D^* = \hat{\beta_D}$*, here* $\hat{\beta_D}$ *is defined implicitly as* $\theta = \frac{\beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\hat{\beta_D}}\right)^2}{\left(\rho_W - \rho_0 + \kappa_D - \frac{1}{4\hat{\beta_D}}\right)}$ *(also note from earlier assumptions* $\hat{\beta}_D > 0$*). The game will proceed as defined in Proposition 1, Case 1.A., where the final out come is the status quo.*

Case 2: $\theta > \frac{\kappa_D}{\rho_W - \rho_0}$

- *2.A. We define* $\check{\beta_D}$ *implicitly as* $\theta = \frac{\kappa_C}{\left(\rho_W - \rho_0 - \frac{\theta}{4\check{\beta_C}} + \frac{1}{2\check{\beta_D}}\right)}$*. As long as* $\frac{\theta}{2\check{\beta_C}} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\check{\beta_D}}$*, then D selects* $\beta_D^* = \check{\beta_D}$*. The game will proceed as defined in Proposition 1, Case 2.C., where the final outcome is gray zone conflict where C is not bound by D's deterrent threat.*
- *2.B. Otherwise, D selects* $\beta_D^* = \dot{\beta_D}$*, here* $\dot{\beta_D}$ *is defined implicitly as* $\theta = \frac{\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\dot{\beta_D}}\right)^2}{-\kappa_D + \frac{1}{4\dot{\beta_D}}}$*. The game will proceed as defined in Proposition 1, Case 1.C., where the final outcome is gray zone conflict where C is not bound by D's deterrent threat.*

As one example of how this one equilibrium plays out, we adapt Figure 4 in the text. Now the solid black lines denote the selected levels of β_D^* (with $1/\beta_D$ plotted so that greater y-axis values represent greater gray zone efficiencies for D), and the dotted lines separate equilibrium spaces.

Moving left to right, for θ between 1.285 and $\frac{\kappa_C}{\rho_W - \rho_0}$, D's optimal β_D^* is described in Proposition A Case 1.A. As the outcome, C will optimally select into the status quo. For this selected β_D^*, C knows that C would face enough of a challenge in gray zone conflict to make competing there too costly. Thus within this region, D could select a low enough β_D^* to compel C to forgo limited challenges and conflict, and stick to the status quo.

Moving right, for θ between $\frac{\kappa_C}{\rho_W - \rho_0}$ and $2\beta_C(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\check{\beta_D}})$, D's optimal β_D^* is described in Proposition A Case 2.A. As the outcome, C will optimally select into gray zone conflict, but will be constrained by C's internal costs. For this selected β_D^*, D wants to challenge C in gray zone conflict (which a lower β_D^* accomplishes), but does not want to push C into forgoing gray zone conflict, because within this region C prefers war to accepting the status quo. Thus here, D selects the β_D^* where C selects into gray zone conflict and is not bound by the deterrent threat, because this gives D some surplus beyond what war or C selecting gray zone conflict and being bound by the deterrent threat produces.

Finally, for θ between $2\beta_C(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\check{\beta_D}})$ and 1.4, D's optimal β_D^* is described in Case 2.B. As the outcome, C will optimally select into gray zone conflict, and will be constrained by D's deterrent threat. This situation is problematic for D. If D modifies β_D^*, either C will adapt by selecting the new g_C^* that makes D indifferent between war and gray zone conflict, or will go to war over the issue. Within this region, it does not matter what β_D^* is selected, because C will always select an action that gives D their wartime utility.

1.5 Extension 2: Probabilistic Escalation to War

A useful feature of the model above is that everything that occurs is deterministic. Only if a state wants to go to war or wants to enter gray zone conflict does it actually happen. However, this represents a simplification. Perhaps in some cases, one state behaving aggressively in lower-levels of conflict can create an incident that necessitates an escalation to higher levels of conflict. To speak to this issue, we introduce the possibility of probabilistic escalation out of gray zone conflict. Our results are substantively similar, but this change shifts some equilibrium properties. Intuitively, now gray zone conflict can probabilistically lead to C's worst outcome: where C invests in limited challenges, war happens, and C must pay the costs of limited challenges with the costs of war. Strategically, because here gray zone conflict is overall worse for C, C will be more willing to accept the status quo or go to war.

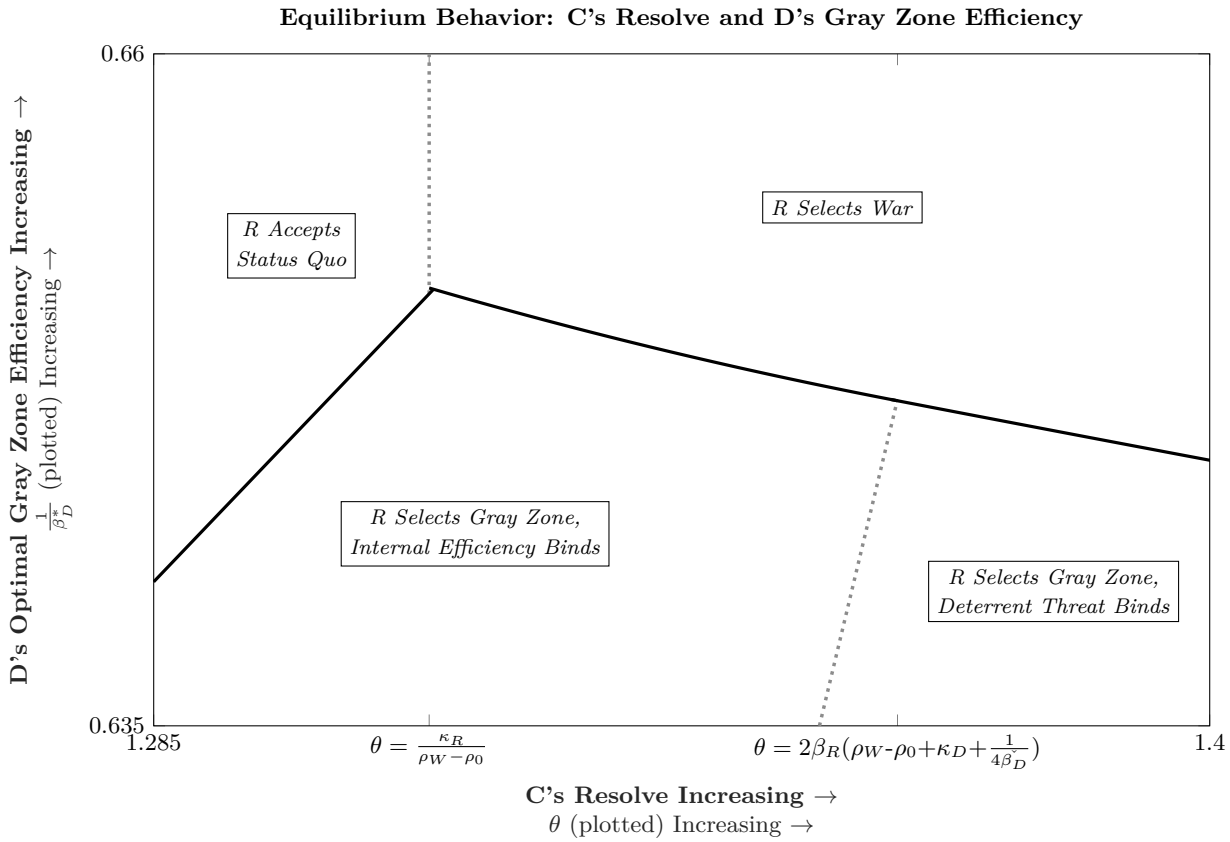

Figure A1: Extension 1: D's Optimal d^*. C's resolve θ and the inverse D's gray zone efficiency $\frac{1}{\beta_D}$ are plotted. The dotted lines separate different kinds of equilibrium play, and the dark black lines denote D's optimal selected β_D. The parameters are $\rho_0 = 0$, $\rho_W = 0.5$, $\beta_C = 1$, $\kappa_C = 0.53$, and $\kappa_D = 0.1$.

There are many possible ways to model this. For ease, we choose (in our opinion) the simplest way, which is that selecting $g_C > 0$ introduces a $1-\zeta \in (0,1)$ likelihood of an escalation to war. Thus, when C selects $g_C > 0$, C's new expected utility is

$$U_C = \theta\left(\zeta P(g_C, g_D) + (1-\zeta)\rho_W\right) - (1-\zeta)\kappa_C - \beta_C g_C.$$

To offer some intuition, g_D^*, $\hat{g_C}$, $\check{g_C}$, and $\tilde{g_C}$ remain the same as it was in the model in the text (as defined in Proposition 1). However, the cut-points that distinguish C's decision to enter into the status quo, gray zone conflict, or war change slightly; overall, the key take-away is that considering probabilistic escalation makes gray zone conflict less appealing relative to the status quo and war.

We express equilibrium behavior in Proposition B. Then below, we derive the new cut-points, Additionally in the derivations, we discuss how the new cut-points imply that gray zone conflict is less appealing and fewer types θ will select into it relative to the game without a probabilistic likelihood of escalation to war from gray zone conflict.

Proposition B: *In equilibrium, the game with a $1-\zeta$ chance of escalation out of gray zone conflict to war will play out in the following manner.*

Case 1, $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$:

- *1.A. If $\theta \leq \frac{(1-\zeta)\kappa_C + \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{(1-\zeta)(\rho_W - \rho_0) + \zeta\left(\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D}\right)}$ and $\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}$, then C accepts the status quo. C selects $w_C^* = 0$ and $g_C^* = 0$, and D selects $w_D^* = 0$ and $g_D^* = 0$.*
- *1.B. If $\theta > \frac{\zeta\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\zeta\left(\frac{1}{4\beta_D} - \kappa_D\right)}$ and $\theta > \frac{\kappa_C}{\rho_W - \rho_0}$, then C declares war. C selects $w_C^* = 1$.*
- *1.C. Otherwise, the game end in gray zone conflict where* C's *limited challenge is constrained by D's deterrent threat. C selects $w_C^* = 0$ and $g_C^* = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$, and (assuming the game does not probabilistically escalate to war) D selects $w_D^* = 0$ and $g_D^* = \frac{1}{2\beta_D}$.*

Case 2, $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$:

- *2.A. If $(1-\zeta)\kappa_C \geq \theta\left((1-\zeta)(\rho_W - \rho_0) + \zeta\left(\frac{\theta}{2\beta_C} - \frac{1}{2\beta_D}\right) - \frac{\theta}{4\beta_C}\right)$ and $\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}$, then C accepts the status quo. C selects $w_C^* = 0$ and $g_C^* = 0$, and D selects $w_D^* = 0$ and $g_D^* = 0$.*
- *2.B. If $\theta > \frac{\zeta\kappa_C}{\left(\zeta\left(\rho_W - \rho_0 - \frac{\theta}{2\beta_C} + \frac{1}{2\beta_D}\right) + \frac{\theta}{4\beta_C}\right)}$ and $\theta > \frac{\kappa_C}{\rho_W - \rho_0}$, then C declares war. C sets $w_C^* = 1$.*[4]
- *2.C. Otherwise, the game will end in gray zone conflict where* C's *limited challenge is constrained by* C's *internal efficiency. C selects $w_C^* = 0$ and $g_C^* = \frac{\theta}{2\beta_C}$, and (assuming the game does not probabilistically escalate to war) D selects $w_D^* = 0$ and $g_D^* = \frac{1}{2\beta_D}$.*

1.5.1 Equilibrium Intuition

First, we consider the case when $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. This implies that C will select $g_C^* = \hat{g_C} = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. We can then express C's behavior in terms of θ. C prefers the status quo to gray zone conflict when

$$\theta\rho_0 \geq \theta\left(\zeta\left(\rho_W + \kappa_D - \frac{1}{4\beta_D}\right) + (1-\zeta)\rho_W\right) - (1-\zeta)\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2$$

or

$$\frac{\beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\zeta\left(\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D}\right)} + \frac{(1-\zeta)(\theta\rho_0 - \theta\rho_W + \kappa_C)}{\zeta\left(\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D}\right)} \geq \theta.$$

[4] *While the right-hand-side of this condition is also increasing in θ, the left-hand-side increases faster with increases in θ.*

Note that the inequality sign does not flip because, by Assumption 1, $\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D} > 0$. We are able to say that $\frac{\beta_C\left(\rho_W-\rho_0+\kappa_D+\frac{1}{4\beta_D}\right)^2}{\zeta\left(\rho_W-\rho_0+\kappa_D-\frac{1}{4\beta_D}\right)} > \frac{\beta_C\left(\rho_W-\rho_0+\kappa_D+\frac{1}{4\beta_D}\right)^2}{\left(\rho_W-\rho_0+\kappa_D-\frac{1}{4\beta_D}\right)}$ because $\zeta \in (0,1)$. Furthermore, this constraint (on when the status quo is preferred to gray zone conflict) matters only when C prefers the status quo to war, or when $\theta\rho_0 - \theta\rho_W + \kappa_C \geq 0$; this condition implies $\frac{(1-\zeta)(\theta\rho_0-\theta\rho_W+\kappa_C)}{\zeta\left(\rho_W-\rho_0+\kappa_D-\frac{1}{4\beta_D}\right)} \geq 0$, which means $\frac{\beta_C\left(\rho_W-\rho_0+\kappa_D+\frac{1}{4\beta_D}\right)^2}{\zeta\left(\rho_W-\rho_0+\kappa_D-\frac{1}{4\beta_D}\right)} + \frac{(1-\zeta)(\theta\rho_0-\theta\rho_W+\kappa_C)}{\zeta\left(\rho_W-\rho_0+\kappa_D-\frac{1}{4\beta_D}\right)} > \frac{\beta_C\left(\rho_W-\rho_0+\kappa_D+\frac{1}{4\beta_D}\right)^2}{\left(\rho_W-\rho_0+\kappa_D-\frac{1}{4\beta_D}\right)}$, which in turn implies that there are more C's with some resolve θ that will select into the status quo in the game here relative to the game in the text without probabilistic escalation.

Next, C prefers war to gray zone conflict when

$$\theta\rho_W - \kappa_C > \theta\left(\zeta\left(\rho_W + \kappa_D - \frac{1}{4\beta_D}\right) + (1-\zeta)\rho_W\right) - (1-\zeta)\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2$$

or

$$\theta > \frac{\zeta\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\zeta\left(\frac{1}{4\beta_D} - \kappa_D\right)}.$$

Note that based on Assumption 1, the above sign does not flip. We can say that $\zeta\kappa_C - \zeta\beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2 > \zeta\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2$. This implies that

$$\frac{\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D} = \frac{\zeta\kappa_C - \zeta\beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\zeta\left(\frac{1}{4\beta_D} - \kappa_D\right)} > \frac{\zeta\kappa_C - \beta_C\left(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\zeta\left(\frac{1}{4\beta_D} - \kappa_D\right)}.$$

In other words, there are more C's with some resolve θ that will select into war in the game here relative to the game without probabilistic escalation.

Next, we assume $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. This condition implies that the selected gray zone conflict will be constrained by C's internal costs and not D's deterrent threat. So, if C selects into gray zone conflict, C will select $g_C^* = \breve{g_C} = \frac{\theta}{2\beta_C}$. We can then express C's behavior in terms of θ. C prefers the status quo to gray zone conflict when

$$\theta\rho_0 \geq \theta\left(\zeta\left(\rho_0 + \frac{\theta}{2\beta_C} - \frac{1}{2\beta_D}\right) + (1-\zeta)\left(\rho_W\right)\right) - (1-\zeta)\kappa_C - \frac{\theta^2}{4\beta_C}$$

or

$$(1-\zeta)\kappa_C \geq \theta\left((1-\zeta)(\rho_W - \rho_0) + \zeta\left(\frac{\theta}{2\beta_C} - \frac{1}{2\beta_D}\right) - \frac{\theta}{4\beta_C}\right).$$

To speak to this inequality, we will need to consider a few different cases here.

First, it could be possible that $\left((1-\zeta)(\rho_W - \rho_0) + \zeta\left(\frac{\theta}{2\beta_C} - \frac{1}{2\beta_D}\right) - \frac{\theta}{4\beta_C}\right) \leq 0$. When this is the case, then C would never want to select into gray zone conflict as doing so would always be strictly worse for C.

Next, consider when $\left((1-\zeta)(\rho_W - \rho_0) + \zeta\left(\frac{\theta}{2\beta_C} - \frac{1}{2\beta_D}\right) - \frac{\theta}{4\beta_C}\right) > 0$ and $(1-\zeta)(\theta\rho_W - \theta\rho_0 - \kappa_C) > 0$. In this case, C's wartime payoff $\theta\rho_W - \kappa_C$ is greater than C's status quo payoff, meaning that C would never select into the status quo over selecting into war, meaning this constraint would never be activated.

Finally, consider when $\left((1-\zeta)(\rho_W-\rho_0)+\zeta\left(\frac{\theta}{2\beta_C}-\frac{1}{2\beta_D}\right)-\frac{\theta}{4\beta_C}\right)>0$ and $(1-\zeta)(\theta\rho_W-\theta\rho_0-\kappa_C)<0$. We can re-write the above as

$$0 \geq \theta\left(\zeta\left(\frac{\theta}{2\beta_C}-\frac{1}{2\beta_D}\right)-\frac{\theta}{4\beta_C}\right)+(1-\zeta)(\theta\rho_W-\theta\rho_0-\kappa_C)$$

Note that $\frac{\theta}{4\beta_C}-\frac{1}{2\beta_D}=\frac{\theta}{2\beta_C}-\frac{1}{2\beta_D}-\frac{\theta}{4\beta_C}>\zeta\left(\frac{\theta}{2\beta_C}-\frac{1}{2\beta_D}\right)-\frac{\theta}{4\beta_C}$, where the inequality holds by Assumption 1. Altogether, this means that $\theta\left(\frac{\theta}{4\beta_C}-\frac{1}{2\beta_D}\right)>\theta\left(\zeta\left(\frac{\theta}{2\beta_C}-\frac{1}{2\beta_D}\right)-\frac{\theta}{4\beta_C}\right)+(1-\zeta)(\theta\rho_W-\theta\rho_0-\kappa_C)$. This implies that there are more C's with some resolve θ that will select into the status quo in the game here relative to the game without probabilistic escalation.

Finally, assuming $\frac{\theta}{2\beta_C}<\rho_W-\rho_0+\kappa_D+\frac{1}{4\beta_D}$, C prefers war to gray zone conflict when

$$\theta\rho_W-\kappa_C>\theta\left(\zeta\left(\rho_0+\frac{\theta}{2\beta_C}-\frac{1}{2\beta_D}\right)+(1-\zeta)\left(\rho_W\right)\right)-(1-\zeta)\kappa_C-\frac{\theta^2}{4\beta_C}$$

or

$$\theta>\frac{\zeta\kappa_C}{\left(\zeta\left(\rho_W-\rho_0-\frac{\theta}{2\beta_C}+\frac{1}{2\beta_D}\right)+\frac{\theta}{4\beta_C}\right)}.$$

Note the inequality sign does not flip because $\left(\rho_W-\rho_0-\frac{\theta}{2\beta_C}+\frac{1}{2\beta_D}\right)>0$. Furthermore, by that condition, $\zeta\left(\rho_W-\rho_0-\frac{\theta}{2\beta_C}+\frac{1}{2\beta_D}\right)+\frac{\theta}{4\beta_C}>\zeta\left(\rho_W-\rho_0-\frac{\theta}{2\beta_C}+\frac{1}{2\beta_D}\right)+\zeta\frac{\theta}{4\beta_C}$. Therefore $\frac{\kappa_C}{\left(\rho_W-\rho_0-\frac{\theta}{2\beta_C}+\frac{1}{2\beta_D}\right)+\frac{\theta}{4\beta_C}}>\frac{\zeta\kappa_C}{\zeta\left(\rho_W-\rho_0-\frac{\theta}{2\beta_C}+\frac{1}{2\beta_D}\right)+\frac{\theta}{4\beta_C}}$. This implies that there are more C's with some resolve θ that will select into war in the game here relative to the game without a random chance of escalation.

Finally, note that D's strategies in this game are unchanged from the game without probabilistic escalation.

1.6 Extension 3: Endogenous Bargaining and Information Asymmetry

Here we offer one possible microfoundation for a key assumption in the game: that the game begins with C being potentially dissatisfied with the status quo. We do this by keeping the game structure we introduced in the main paper and by adding new initial moves to the game. We grant D the option to establish the "status quo" though an ultimatum offer, we assume D has private information, and we add some additional parameter assumptions. Importantly, reasonable readers may take issue with certain facets of the game form below; for that reason, we wish to highlight that what we present is not the only way for our conflict selection subgame (i.e. C selecting whether to accept, go to war, or engage in a low level challenge, and then D doing the same) to start with a potentially dissatisfied C. Past examinations of the causes for war—like private information (Fearon 1995), commitment problems (Powell 2006), costly peace (Coe 2011), political bias (Jackson and Morelli 2007), and the possibility of behavioral types (Acharya and Grillo 2015)—have all demonstrated that natural circumstances can lead to settings where one state is willing to make an offer to another state, knowing that the offer could lead to some form of conflict.

The take-away from the following analysis is that with endogenous bargaining, C's gray zone actions are still driven by the external deterrent constraint and the internal efficiency constraint, though other factors can also play a role. While it is well established that including information asymmetry can make an empirical analysis of conflict onset less pinned-down (Gartzke 1999), the model does predict that whenever gray zone conflict is observed, C will select a level of gray zone conflict benchmarked either to the external deterrent constraint of D, or based on their own internal efficiency constraint. Based on this prediction, we also include an empirical analysis in A7 below where we re-run our models on the sample of observations where some form of conflict occurred. We find that our key variables approximating the deterrent threat (NATO membership) and internal efficiency (distance from Russia) still predict the intensity of conflict. While these results should

be taken with precautions—the sample is smaller and (due to the high number of recent attacks) a larger proportion of control variables are missing—this does still suggest our formal model is plausibly describing key drivers for gray zone conflict intensity.

For this new formal model, we assume the following game form.

1. Nature moves first and sets $\rho_W \in \{\underline{\rho}_W, \bar{\rho}_W\}$, with $\underline{\rho}_W < \bar{\rho}_W$. Nature fixes $\bar{\rho}_W$ with probability $\epsilon \in (0,1)$ and $\underline{\rho}_W$ with probability $1-\epsilon$. D observes the selected ρ_W, but C does not. The selected ρ_W can be thought of as D's "type," where D is stronger if they are type $\underline{\rho}_W$.
2. D makes C some offer $x \in \{\underline{x}, \bar{x}\}$, with $\underline{x} < \bar{x}$. It is worthwhile mentioning that issue indivisibility is not a driver of conflict in this game, and if we assumed complete information, there will be no conflict.[5]
3. C either goes to war by setting $w_C = 1$ or sets $w_C = 0$. If C goes to war, the game terminates and C and D receive payoffs $\theta\rho_W - \kappa_C$ and $1 - \rho_W - \kappa_D$ (with $\rho_W \in \{\underline{\rho}_W, \bar{\rho}_W\}$), respectively. If C sets $w_C = 0$, C also selects $g_C \in \mathcal{G}_C = \mathbb{R}_{\geq 0}$, where $g_C = 0$ is walking away from the crisis and accepting the offer, and $g_C > 0$ is conducting some limited, costly military action that shifts the offer in favor of the challenger.
4. As long as C did not previously go to war, D can either escalate to war by setting $w_D = 1$, or not by setting $w_D = 0$ and selecting some gray zone response $g_D \in \mathcal{G}_D = \mathbb{R}_{\geq 0}$, with $g_D = 0$ implying that D does not respond to the limited challenge. When war occurs, C and D receive payoffs $\theta\rho_W - \kappa_C$ and $1 - \rho_W - \kappa_D$. When D selects $g_D \geq 0$, C and D receive payoffs $\theta P(x, g_C, g_D) - \beta_C g_C^2$ and $1 - P(x, g_C, g_D) - \beta_D g_D^2$, with $P(x, g_C, g_D) = max\{min\{\bar{\rho}_W, x + g_C - g_D\}, x\}$.

The payoffs to the game are summarized in the table below. There are two key changes to highlight here.

First, we modify C's utility should D escalate to war after C selects some level of gray zone conflict. Here C faces no costs from this gray zone challenge. This assumption is made primarily for analytic ease. In the mixing equilibrium examined below, under some parameters, C may select a limited challenge that provokes strong-type D's to go to war. When gray zone challenge costs do not "carry-over" (as is assumed in this model) C selects their optimal gray zone challenge purely based on the weak-type D's parameters. If C faces carry-over costs and the carry-over costs of gray zone challenges when D declares war are too high, C may reduce their optimal gray zone challenge in order to mitigate some carry-over costs when C is paired with a strong-type D.[6] So long that the carry-over costs are sufficiently low, sufficiently dampened, or that D is a strong-type with a low-enough probability, this assumption makes little difference.[7]

Second, the P function now falls between the offer x and the expected political war outcome for the weak-type D. This still embraces that gray zone conflict is a limited challenge.

Scenario	**C's utility facing $\underline{\rho}_W$**	**C's utility facing $\bar{\rho}_W$**	**D's utility ($\rho_W \in \{\underline{\rho}_W, \bar{\rho}_W\}$)**
C initially initiates war ($w_C = 0$)	$\theta\underline{\rho}_W - \kappa_C$	$\theta\bar{\rho}_W - \kappa_C$	$1 - \rho_W - \kappa_D$
C and D select gray zone/accept status quo ($w_C = 0$, $g_C \geq 0$, $w_D = 0$, $g_D \geq 0$)	$\theta P(x, g_C, g_D) - \beta_C g_C^2$	$\theta P(x, g_C, g_D) - \beta_C g_C^2$	$1 - P(x, g_C, g_D) - \beta_D g_D^2$
D escalates to war after C acts ($w_C = 0$, $g_C \geq 0$, $w_D = 1$)	$\theta\underline{\rho}_W - \kappa_C$	$\theta\bar{\rho}_W - \kappa_C$	$1 - \rho_W - \kappa_D$

Table A1: Summarized payoffs for actors

We now examine perfect Bayesian Nash equilibria. To accommodate additional types, we slightly adapt Assumptions 1 from above. This is now the following.

***Assumption 1C**: In equilibrium, $x < P(\tilde{g_C}, \tilde{g_D}) < \bar{\rho}_W$, where $x \in \{\underline{x}, \bar{x}\}$ is the selected offer.*

[5] This model can still function with a continuum of offers, though it becomes much more complicated.

[6] If this assumption were not in place, this would change C's selected gray zone challenge cases 1D and 2D within Proposition C, and it would alter the decisions over which case to enter; this is what was observed in the second extension.

[7] As the most reasonable way to accommodate the assumptions in the paper and here, assume that the carry-over costs are non-zero but very small.

Based on the optimal $\tilde{g_C}$ and $\tilde{g_D}$ (solved below), for a given x, this condition amounts to $\frac{\theta}{2\beta_C} - \frac{1}{2\beta_D} > 0$ and $0 < \bar{\rho}_W - x - \frac{\theta}{2\beta_C} + \frac{1}{2\beta_D}$ if $\frac{\theta}{2\beta_C} < \bar{\rho}_W - x + \kappa_D + \frac{1}{4\beta_D}$, and $\bar{\rho}_W - x + \kappa_D - \frac{1}{4\beta_D} > 0$ and $\kappa_D - \frac{1}{4\beta_D} < 0$ if $\bar{\rho}_W - x + \kappa_D + \frac{1}{4\beta_D} \leq \frac{\theta}{2\beta_C}$.

We also make several assumptions on $\underline{x}$ and $\bar{x}$. First, we assume that if C selects any kind of gray zone challenge after receiving an offer of $\underline{x}$, a strong-type D ($\underline{\rho}_W$) will go to war. This is the following, but notably, there are other ways for this assumption to hold as well (Abreu and Gul 2000; Acharya and Grillo 2015).[8]

Assumption 2C: $\underline{x} = \underline{\rho}_W + \kappa_D$.

We also assume that weak-type D's ($\bar{\rho}_W$) prefer attaining a final payoff of $1 - \bar{x}$ (i.e. making a high-offer to C) rather than setting $x = \underline{x}$ and experiencing gray zone conflict or war. This produces:

Assumption 3C: $1 - \bar{x} > 1 - \bar{\rho}_W - \kappa_D$, and if $\frac{\theta}{2\beta_C} < \bar{\rho}_W - \bar{x} + \kappa_D + \frac{1}{4\beta_D}$ then $1 - \bar{x} > 1 - \underline{x} - \frac{\theta}{2\beta_C} + \frac{1}{4\beta_D}$.

We also assume that if C receives a high offer ($\bar{x}$) from a weak-type D, C will accept the offer rather than escalate to war or implement gray zone conflict. This produces:

Assumption 4C: $\theta \leq \frac{\kappa_C}{\bar{\rho}_W - \bar{x}}$ and if $\frac{\theta}{2\beta_C} \geq \bar{\rho}_W - \bar{x} + \kappa_D + \frac{1}{4\beta_D}$ then $\theta \leq \frac{\beta_C\left(\bar{\rho}_W - \bar{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\bar{\rho}_W - \bar{x} + \kappa_D - \frac{1}{4\beta_D}\right)}$, or if $\frac{\theta}{2\beta_C} < \bar{\rho}_W - \bar{x} + \kappa_D + \frac{1}{4\beta_D}$ then $0 \geq \theta\left(\frac{\theta}{4\beta_C} - \frac{1}{2\beta_D}\right)$.

And finally, we assume that if C receives a low offer ($\underline{x}$) from a strong type D, then C prefers accepting the offer rather than going to war.

Assumption 5C: $\underline{x} > \underline{\rho}_W - \kappa_C$.

With these assumptions in place, the complete information game would play out as follows. Strong-type D's (when nature sets $\underline{\rho}_W$) would always make low offer $\underline{x}$ to C, and C would always accept (based on Assumptions 2C and 5C). Weak-type D's (when nature sets $\bar{\rho}_W$) choose one of two offers. First, suppose that in response to a low offer of $\underline{x}$ by a type $\bar{\rho}_W$, C's best response is to accept. Formally, this would imply that, when facing a weak-type D, C prefers accepting to war ($\underline{x} > \theta\bar{\rho}_W - \kappa_C$) and accepting to gray zone conflict (if $\frac{\theta}{2\beta_C} \geq \bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}$ then $\theta \leq \frac{\beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\bar{\rho}_W - \underline{x} + \kappa_D - \frac{1}{4\beta_D}\right)}$, and if $\frac{\theta}{2\beta_C} \geq \bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}$ then $0 \geq \theta\left(\frac{\theta}{4\beta_C} - \frac{1}{2\beta_D}\right)$). If this is the case, then weak-type D's will pool with strong-type D's, also set $x = \underline{x}$, and C will always accept the low offer. Second, suppose that in in response to a low-offer of $\underline{x}$ by a type $\bar{\rho}_W$, C's best response is to engage in a limited challenge or war. If this is the case, then weak-type D will always make a high offer to C rather than face any kind of conflict (based on Assumptions 3C and 4C). Thus, in the complete information version of the game, under the above assumptions, there will be no war or gray zone conflict.

Next, consider what happens when information asymmetry is introduced. Strong-type D's ($\underline{\rho}_W$) will always make the low offer $\underline{x}$ (based on Assumptions 2C and 5C). A fully separating equilibrium is not possible,[9] and type $\bar{\rho}_W$ D's will pool on $\underline{x}$ or semi-pool by mixing between the $\underline{x}$ and $\bar{x}$ offers. In response to a high offer of $\bar{x}$, C will accept ($g_C = 0$) by virtue of Assumption 4C. In response to a low offer of $\underline{x}$, C may mix, sometimes accepting ($g_C = 0$) and sometimes selecting some optimal response similar to what is outlined in Proposition 1 in the paper (with the new full equilibrium outlined below in Proposition C). Essentially, when a weak-type D makes the low offer ($\underline{x}$) to C, C's behavior is very close to the "dissatisfied state" discussed in the paper, who has been presented with a political status quo (the offer) that could tempt them to go to war.

We derive the equilibria here. For a semi-separating equilibrium to exist, weak-type D's must mix between making a high offer ($\bar{x}$) to C that will result in peace, and a low offer $\underline{x}$ to C, knowing that C will, with

[8] The simplest way to implement this would be to assume that the strong-type D is a behavioral type that will go to war if facing any challenge.

[9] Suppose a separating equilibrium exists where strong type D's select $x = \underline{x}$ and weak type D's select $x = \bar{x}$. C does not want to go to war with type $\underline{\rho}_W$ D's and would therefore never challenge or go to war when facing low offer $x = \underline{x}$. However, this would incentivize weak-type D's to deviate to $x = \underline{x}$.

probability $\alpha \in (0,1)$, accept the low offer. What C does otherwise (with probability $1-\alpha$) is solved for below. For now, call what C does otherwise (with probability $1-\alpha$) C's "conflict response." Assume for now that a type $\bar{\rho}_W$ facing the conflict response after making the low offer $\underline{x}$ does worse than the type $\bar{\rho}_W$ D would have done by making the high offer $\bar{x}$.[10] Formally, letting $U_D(\bar{\rho}_W, CR)$ denote type $\bar{\rho}_W$ D's utility from C's conflict response (CR), this implies $U_D(\bar{\rho}_W, CR) < 1-\bar{x}$.

We solve for α below. For weak type $\bar{\rho}_W$ D's to be indifferent between making the high and low offer, the following condition must hold:

$$1-\bar{x} = \alpha(1-\underline{x}) + (1-\alpha) * U_D(\bar{\rho}_W, CR),$$

or, in terms of α,

$$\alpha = \frac{1-\bar{x}-U_D(\bar{\rho}_W, CR)}{1-\underline{x}-U_D(\bar{\rho}_W, CR)}.$$

We will reference this term again in the statement of the equilibria. Note that because $\bar{x} > \underline{x}$ and Assumption 3C, α always falls within 0 and 1 (non-inclusive).

Next, we assume that with probability $\gamma \in (0,1)$ that type $\bar{\rho}_W$ D's will make low offer $\underline{x}$. We use Bayes' rule to calculate probabilities of type conditional on offer. These are

$$Pr(\bar{\rho}_W|\underline{x}) = \frac{Pr(\underline{x}|\bar{\rho}_W) * Pr(\bar{\rho}_W)}{Pr(\underline{x})}$$

or

$$Pr(\bar{\rho}_W|\underline{x}) = \frac{(1-\epsilon)\gamma}{\epsilon + (1-\epsilon)\gamma},$$

and

$$Pr(\underline{\rho}_W|\underline{x}) = \frac{\epsilon}{\epsilon + (1-\epsilon)\gamma}.$$

This next derivation also depends on C's conflict response. We assume C attains utility $U_C(\bar{\rho}_W, CR)$ when selecting their optimal conflict response and facing a type $\bar{\rho}_W$ D. C is indifferent between their conflict response and accepting low offer $\underline{x}$ when

$$\underline{x} = \frac{(1-\epsilon)\gamma}{\epsilon + (1-\epsilon)\gamma} U_C(\bar{\rho}_W, CR) + \frac{\epsilon}{\epsilon + (1-\epsilon)\gamma}\left(\underline{\rho}_W - \kappa_C\right)$$

or

$$\gamma = \frac{\epsilon\left(\underline{x} - \left(\underline{\rho}_W - \kappa_C\right)\right)}{(1-\epsilon)\left(U_C(\bar{\rho}_W, CR) - \underline{x}\right)}.$$

The set of derivations above are all for semi-separating equilibria. In Proposition C below, Cases 1B, 1D, 2B, and 2D are all semi-separating equilibria following the structure described above.

Pooling equilibria could also exist. For example, suppose that C's optimal conflict response to a type $\bar{\rho}_W$ D offering $\underline{x}$ is to accept the bargained offer ($w_C = 0$ and $g_C = 0$). If this is the case, then type $\bar{\rho}_W$ D's does best by always fixing $x = \underline{x}$ because C cannot credibly commit to any form of conflict. In Proposition C below, Cases 1A and 2A take this form. These equilibria can be thought of as some combination of D's external deterrent threat and C's internal efficiency constraints binding. Essentially C is so ineffective at fighting war and gray zone conflict that C does best walking away with a low-offer rather than fighting.

[10] This is partly justified by Assumption 3C; however, sometimes C's optimal conflict response is accepting the status quo.

Another type of pooling equilibrium can also exist. Sometimes C's war outcome against strong-type D's that C would never risk challenging a low-offer ($\underline{x}$) even if all type $\bar{\rho}_W$ D's (with probability $\gamma = 1$) are making the low-offer. Cases 1C, 1E, 2C, and 2E take this form. In these cases, C's optimal conflict response to a type $\bar{\rho}_W$ D offering $\underline{x}$ is some gray zone challenge or war, but C's expected utility from this optimal conflict response is too low for C to consider challenging due to the presence of strong-type D's. These equilibria are certainly shaped by D's external deterrent threat (as the deterrent threat from war with type $\underline{\rho}_W$ D's is central), and they may also be shaped by C's internal efficiency constraints; if C is ineffective at gray zone conflict, then challenging type $\bar{\rho}_W$ D's is less appealing.

The behavior that is described in Proposition 1 (in the text) is related to cases 1B, 1D, 2B, and 2D in Proposition C whenever a weak-type D makes a low-offer and C selects their optimal conflict response. Similarly, in cases 1A and 2A, the equilibria behavior matches what is in Proposition 1. In all of these cases, C is tailoring their gray zone challenge (or no challenge) based on weak-type D's deterrent threat or their own internal efficiency constraint.

Admittedly, there are several differences. For one, even within cases 1B, 1D, 2B, and 2D in Proposition C, the presences of the strong-type D's and the mixed strategies makes these equilibria play out differently. Additionally, the conditions for selection into the various equilibria have been refined further. What the proposition below suggests is that there is further nuance to D's deterrent threat and C's internal efficiency constraint when considering environments with information asymmetry. We can now write out the full equilibria conditions.

However, a key take-away from the equilibrium below is that, conditional on C selecting into gray zone conflict, C will make this selection based on its own internal efficiency constraint or on D's external deterrent constraint, as demonstrated in 1D and 2D.

Proposition C: *In equilibrium, the game will play out in the following manner.*

Case 1, $\frac{\theta}{2\beta_C} \geq \bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}$*:*

- *1.A. If* $\theta \leq \frac{\beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\bar{\rho}_W - \underline{x} + \kappa_D - \frac{1}{4\beta_D}\right)}$ *and* $\theta \leq \frac{\kappa_C}{\bar{\rho}_W - \underline{x}}$*, then both types of D always offer* $x^* = \underline{x}$ *and C always accepts the offer. C selects* $w_R^* = 0$ *and* $g_C^* = 0$*, and both types of D select* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*.*

- *1.B. If* $\theta > \frac{\kappa_C - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D}$*,* $\theta > \frac{\kappa_C}{\bar{\rho}_W - \underline{x}}$*, and* $\theta\underline{x} < \epsilon\left(\theta\bar{\rho}_W - \kappa_C\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$ *then type* $\underline{\rho}_W$ *D always offers* $x^* = \underline{x}$ *and type* $\bar{\rho}_W$ *D offers* $x^* = \underline{x}$ *with probability* γ *and offers* $x^* = \bar{x}$ *with probability* $1 - \gamma$*. When the offer is* $x^* = \bar{x}$*, C always accepts the offer by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, the type* $\bar{\rho}_W$ *D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \bar{x}$ *and* $U_C = \theta\bar{x}$*. When the offer is* $x^* = \underline{x}$*, C accepts the offer with probability* α *by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, in response the both type D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*. When the offer is* $x^* = \underline{x}$*, with probability* $1 - \alpha$ *C declares war; C selects* $w_R^* = 1$*, and payoffs are* $U_D = 1 - \bar{\rho}_W - \kappa_D$ *and* $U_C = \theta\bar{\rho}_W - \kappa_C$ *when D is type* $\bar{\rho}_W$*, and* $U_D = 1 - \underline{\rho}_W - \kappa_D$ *and* $U_C = \theta\underline{\rho}_W - \kappa_C$ *when D is type* $\underline{\rho}_W$*. The values* γ *and* α *are derived using values* $U_D(\bar{\rho}_W, CR) = 1 - \bar{\rho}_W - \kappa_D$ *and* $U_C(\bar{\rho}_W, CR) = \theta\bar{\rho}_W - \kappa_C$*.*

- 1.C. *If* $\theta > \frac{\kappa_C - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D}$*,* $\theta > \frac{\kappa_C}{\bar{\rho}_W - \underline{x}}$*, and* $\theta\underline{x} \geq \epsilon\left(\theta\bar{\rho}_W - \kappa_C\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$*, then both types of D always offer* $x^* = \underline{x}$ *and C always accepts the offer. C selects* $w_R^* = 0$ *and* $g_C^* = 0$*, and both types of D select* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*.*

- *1.D. If* $\theta > \frac{\beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\bar{\rho}_W - \underline{x} + \kappa_D - \frac{1}{4\beta_D}\right)}$*,* $\theta \leq \frac{\kappa_C - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D}$*, and*

  $\theta\underline{x} < \epsilon\left(\theta\left(\bar{\rho}_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$ *, then type* $\underline{\rho}_W$ *D always offers* $x^* = \underline{x}$ *and type* $\bar{\rho}_W$ *D offers* $x^* = \underline{x}$ *with probability* γ *and offers* $x^* = \bar{x}$ *with probability* $1 - \gamma$*. When the offer is* $x^* = \bar{x}$*, C always accepts the offer by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, the type* $\bar{\rho}_W$ *D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \bar{x}$ *and* $U_C = \theta\bar{x}$*. When the offer*

is $x^* = \underline{x}$*, C accepts the offer with probability* α *by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, in response the both type D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*. When the offer is* $x^* = \underline{x}$*, with probability* $1 - \alpha$*, C conducts a limited challenge that is constrained by type* $\bar{\rho}_W$ *D's deterrence threat. C selects* $w_R^* = 0$ *and* $g_C^* = \bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}$*, and in response type* $\underline{\rho}_W$ *D's declare war setting* $w_D^* = 1$*and type* $\bar{\rho}_W$ *D's select* $w_D^* = 0$ *and* $g_D^* = \frac{1}{2\beta_D}$*. The payoffs are* $U_C = \theta\left(\bar{\rho}_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2$ *and* $U_D = 1 - \bar{\rho}_W - \kappa_D$ *when D is type* $\bar{\rho}_W$*, and* $U_C = \theta\underline{\rho}_W - \kappa_C$ *and* $U_D = 1 - \underline{\rho}_W - \kappa_D$ *when D is type* $\underline{\rho}_W$*. The values* γ *and* α *are derived using values* $U_D(\bar{\rho}_W, CR) = 1 - \bar{\rho}_W - \kappa_D$ *and* $U_C(\bar{\rho}_W, CR) = \theta\left(\bar{\rho}_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2$*.*

- *1.E. If* $\theta > \frac{\beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\left(\bar{\rho}_W - \underline{x} + \kappa_D - \frac{1}{4\beta_D}\right)}$*,* $\theta \leq \frac{\kappa_C - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2}{\frac{1}{4\beta_D} - \kappa_D}$*, and*

  $\theta\underline{x} \geq \epsilon\left(\theta\left(\bar{\rho}_W + \kappa_D - \frac{1}{4\beta_D}\right) - \beta_C\left(\bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}\right)^2\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$*, then both types of D always offer* $x^* = \underline{x}$ *and C always accepts the offer. C selects* $w_R^* = 0$ *and* $g_C^* = 0$*, and both types of D select* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*.*

Case 2, $\frac{\theta}{2\beta_C} < \bar{\rho}_W - \underline{x} + \kappa_D + \frac{1}{4\beta_D}$*:*

- *2.A. If* $\theta \leq \frac{2\beta_C}{\beta_D}$ *and* $\theta \leq \frac{\kappa_C}{\bar{\rho}_W - \underline{x}}$*, then both types of D always offer* $x^* = \underline{x}$ *and C always accepts the status quo. C selects* $w_R^* = 0$ *and* $g_C^* = 0$*, and both types of D select* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*.*

- *2.B. If* $\theta > \frac{\kappa_C}{\bar{\rho}_W - \underline{x} - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}$*,* $\theta > \frac{\kappa_C}{\bar{\rho}_W - \underline{x}}$*, and* $\theta\underline{x} < \epsilon\left(\theta\bar{\rho}_W - \kappa_C\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$*, then type* $\underline{\rho}_W$ *D always offers* $x^* = \underline{x}$ *and type* $\bar{\rho}_W$ *D offers* $x^* = \underline{x}$ *with probability* γ *and offer* $x^* = \bar{x}$ *with probability* $1 - \gamma$*. When the offer is* $x^* = \bar{x}$*, C always accepts the offer by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, the type* $\bar{\rho}_W$ *D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \bar{x}$ *and* $U_C = \theta\bar{x}$*. When the offer is* $x^* = \underline{x}$*, with probability* α *C accepts the offer by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, in response the both type D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*. When the offer is* $x^* = \underline{x}$*, with probability* $1 - \alpha$ *C declares war setting* $w_R^* = 1$*, and payoffs are* $U_D = 1 - \bar{\rho}_W - \kappa_D$ *and* $U_C = \theta\bar{\rho}_W - \kappa_C$ *when D is type* $\bar{\rho}_W$*, and* $U_D = 1 - \underline{\rho}_W - \kappa_D$ *and* $U_C = \theta\underline{\rho}_W - \kappa_C$ *when D is type* $\underline{\rho}_W$*. The values* γ *and* α *are derived using values* $U_D(\bar{\rho}_W, CR) = 1 - \bar{\rho}_W - \kappa_D$ *and* $U_C(\bar{\rho}_W, CR) = \theta\bar{\rho}_W - \kappa_C$*.*

- *2.C. If* $\theta > \frac{\kappa_C}{\bar{\rho}_W - \underline{x} - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}$*,* $\theta > \frac{\kappa_C}{\bar{\rho}_W - \underline{x}}$*, and* $\theta\underline{x} \geq \epsilon\left(\theta\bar{\rho}_W - \kappa_C\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$*, then both types of D always offer* $x^* = \underline{x}$ *and C always accepts the status quo. C selects* $w_R^* = 0$ *and* $g_C^* = 0$*, and both types of D select* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*.*

- *2.D. If* $\theta > \frac{2\beta_C}{\beta_D}$*,* $\theta \leq \frac{\kappa_C}{\bar{\rho}_W - \underline{x} - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}$*,* $1 - \underline{x} - \frac{\theta}{2\beta_C} + \frac{1}{4\beta_D} < 1 - \bar{x}$*, and* $\theta\underline{x} < \epsilon\left(\theta\underline{x} + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}\right) + (1-\epsilon)\left(\theta\underline{\rho}_W - \kappa_C\right)$*, then type* $\underline{\rho}_W$ *D always offers* $x = \underline{x}$ *and type* $\bar{\rho}_W$ *D offers* $x = \underline{x}$ *with probability* γ *and offer* $x = \bar{x}$ *with probability* $1 - \gamma$*. When the offer is* $x = \bar{x}$*, C always accepts the offer by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, the type* $\bar{\rho}_W$ *D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \bar{x}$ *and* $U_C = \theta\bar{x}$*. When the offer is* $x = \underline{x}$*, C accepts the offer with probability* α *by setting* $w_R^* = 0$ *and* $g_C^* = 0$*, in response the both type D's set* $w_D^* = 0$ *and* $g_D^* = 0$*, and payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta\underline{x}$*. When the offer is* $x = \underline{x}$*, with probability* $1 - \alpha$ *C conducts a limited challenge that is constrained by C's own internal cost constraints. C selects* $w_R^* = 0$ *and* $g_C^* = \frac{\theta}{2\beta_C}$*, and in response type* $\underline{\rho}_W$ *D's declare war* $(w_D = 1)$ *and type* $\bar{\rho}_W$ *D's select* $w_D^* = 0$ *and* $g_D^* = \frac{1}{2\beta_D}$*. The payoffs are* $U_C = \theta\underline{x} + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}$ *and* $U_D = 1 - \rho_0 - \frac{\theta}{2\beta_C} + \frac{1}{4\beta_D}$ *when D is type* $\bar{\rho}_W$*, and* $U_C = \theta\underline{\rho}_W - \kappa_C$ *and* $U_D = 1 - \underline{\rho}_W - \kappa_D$ *when D is type* $\underline{\rho}_W$*. The values* γ *and* α *are derived using values* $U_D(\bar{\rho}_W, CR) = 1 - \underline{x} - \frac{\theta}{2\beta_C} + \frac{1}{4\beta_D}$*, and* $U_C(\bar{\rho}_W, CR) = \theta\underline{x} + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}$*.*

- *2.E. If* $\theta > \frac{2\beta_C}{\beta_D}$*,* $\theta \leq \frac{\kappa_C}{\bar{\rho}_W - \underline{x} - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}$*,* $1 - \underline{x} - \frac{\theta}{2\beta_C} + \frac{1}{4\beta_D} < 1 - \bar{x}$*, and* $\theta\underline{x} \geq \epsilon\left(\theta\underline{x} + \frac{\theta^2}{4\beta_C} - \frac{\theta}{2\beta_D}\right) + (1 -$

$\epsilon) \left(\theta \underline{\rho}_W - \kappa_C\right)$, *then both types of D always offer* $x = \underline{x}$ *and C always accepts the status quo. C selects* $w_R^* = 0$ *and* $g_C^* = 0$*, and both types of D select* $w_D^* = 0$ *and* $g_D^* = 0$*. Payoffs are* $U_D = 1 - \underline{x}$ *and* $U_C = \theta \underline{x}$*.*

2 New data

The universe of cases was created by first identifying cases of Russian foreign interventions from 3 prior datasets; ICB (Brecher and Wilkenfeld 1997), DCID (Valeriano and Maness 2014), and REI (Casey and Way 2017). Code replicating those findings is provided in the appropriate RMarkdown files. These cases were then supplemented with additional cases of Russian interference the authors were able to identify.

2.1 Comparison of current datasets

A comparison of what cases were covered in each individual dataset is provided in Figure A2. Note that there are significant inconsistencies concerning the sample of post-1994 Russian interventions identified by the ICB, DCID, and REI datasets.

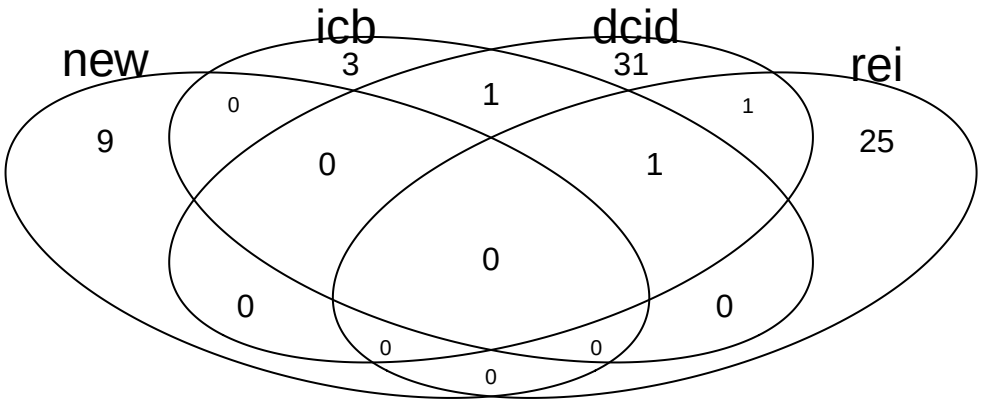

Figure A2: Venn diagram of case overlap among prior datasets

Aside from the cases covered, the intensity codings for current datasets are difficult to compare given their different scales. A more thorough analysis is provided in the appropriate R Markdown files, but a comparison of intensity codings in DCID (Valeriano and Maness 2014) and REI (Casey and Way 2017) is depicted in Figure A3. The DCID data identifies the United States, United Kingdom, Poland and Ukraine as targets of the most severe Russian cyber operations. In the cases documented by REI, the most severe Russian attacks occurred against France, Austria, and Ukraine. Part of this discrepancy is due to the respective foci of each dataset; DCID seeks out cases of cyber incidents and disputes while REI focuses on Russian electoral interference. While a majority of the REI cases include some form of Russian cyber activity, there are a few cases where only material support was provided (eg. Moldova 2014 and Belarus 1994).

This discrepancy exemplifies not only the challenges of relying on open source reporting for identifying cyber influence or disruption campaigns, but also differences in defining what counts as an attack. The only country-year that appears in both datasets is Ukraine 2014. We standardized codings across the two datasets using variable definitions from respective codebooks. A severity less than or equal to 2 in DCID's coding is synonymous in our recoding with REI's coding for disinformation, a severity between 3 and 7 equals REI's coding for cyberattack, and no cases in DCID have a severity greater than 7. We adopted Valeriano and Maness (2014)'s approach of sampling on intensity when there are multiple observations in a given time unit.

2.2 Variable codings

For each incident, we code whether Russia used conventional ground forces, conventional air or sea forces, paramilitary or covert forces, cyber disruption, and information operations. By distinguishing between these

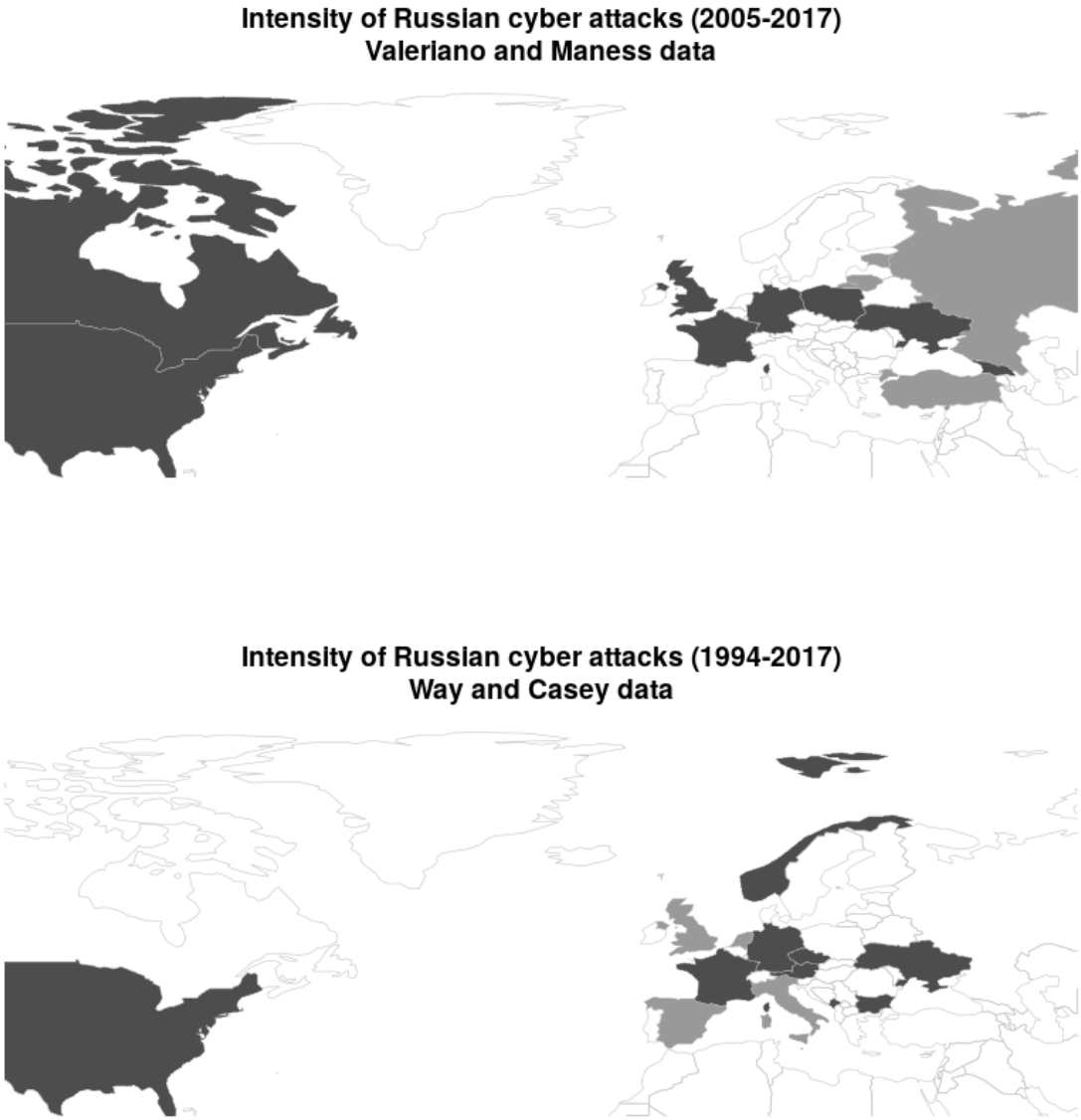

Figure A3: Comparison of coding for highest intensity Russian intervention in each target state

five types of aggression, we obtain a clearer picture of the intensity of each case of Russian intervention. The vast majority of cases include at least some type of cyber operations. In a few cases, data limitations preclude coding of non-kinetic activity by Russia or other actors. In Moldova 2005, for example, Russia provided material support for the Communist Party but there is no credible evidence of cyber activities.

The following binary coding criteria were used for each case:

- **resp_infoops** - Did Russia use information operations during this event? That includes propaganda, misinformation campaigns, theft of information, and other simple intrusions
- **resp_cyberdisrup** - Did Russia use cyber attacks during this operation? That includes hacking, phishing, cyber espionage, DDoS attacks, etc. that constitute a system shut down rather than simple

intrusions

- **resp_paramil** - Did Russia use paramilitary troops during this event? Special forces, covert troops, speznatz, etc all count
- **resp_convmil_airsea** - Did Russia use conventional naval or air forces during this event?
- **resp_convmil_gro** - Did Russia use conventional ground troops like their army, artillery, tanks, etc during this event?

The complete dataset is provided in the appropriate .csv file. It includes sources used for the codings as well as justifications and explanations where needed.

2.3 Summary statistics

Although data was compiled on Russian intervention against all states from 1994-2018, the statistical analysis is limited to a sample from European states. In alignment with that, Table A2 present descriptive statistics of the sample used in the models provided in the main text.

Table A2: Covariate Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Intensity	1,000	0.1	0.4	0	0	0	5
NATO member	1,000	0.5	0.5	0	0	1	1
Dist. from Russia (minimum, log)	1,000	5.2	2.9	0.01	5.2	7.0	7.8
Democracy	926	0.9	0.3	0.0	1.0	1.0	1.0
Nuclear state	1,000	0.05	0.2	0	0	0	1
Population (log)	1,000	15.8	1.4	12.5	14.9	16.3	18.2
CINC ratio	754	0.1	0.1	0.0	0.01	0.1	0.4
GDP per capita	995	26.6	23.8	0.7	6.7	41.4	112.0
Military expenditure	962	7.3	13.3	0.0	0.3	5.7	59.8

Sample includes all European states (1994-2018). Binary variables converted to numeric.

The distribution of our dependent variable, intensity, is shown for the European sample in Figure A4. The figure only includes country-years with known attacks (omitting null cases) to allow an easier visual comparison of variation in attack intensity.

The bivariate correlations between the DV and the two EVs are shown in Figure A5. The intensity and NATO variables have been converted to numeric values to simplify visualizing the bivariate correlations.

3 Alternate model specifications

We run a set of alternate model specifications as robustness checks. Our results are consistent across alternate modeling specifications including different regression models, control variables, and imputation strategies. We choose the ordered probit results as the main results given the appropriateness of that model specification and to ensure our primary results are not simply an artifact of our imputation strategy. Those results are shown below.

3.1 Alternate alliance measure

NATO membership is a crude observable measure of alliances that proxy for deterrence. In **??** use an alternate measure that also includes pre-NATO membership, which accounts for country-years in the PfP, MAP, or Intensified Dialogue stage regardless of contemporary NATO status (meaning a state in MAP at the time of this writing is considered pre-NATO even if it has not yet joined NATO). We run the new models using all models from both samples in the main text.

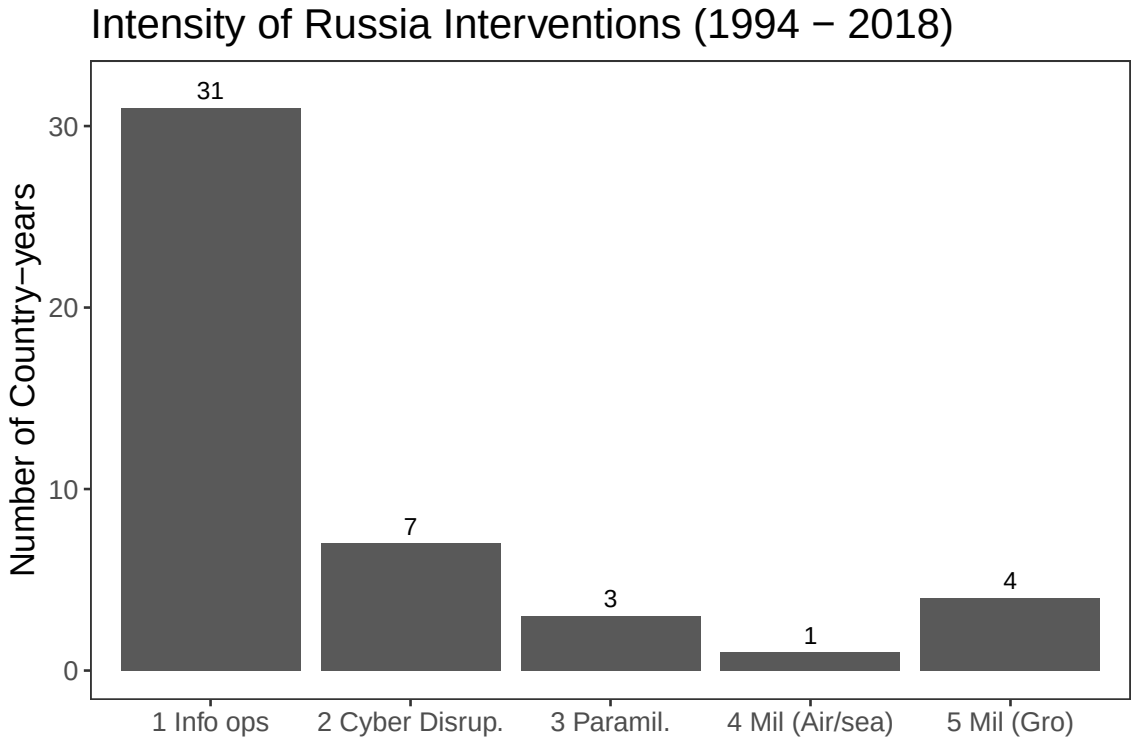

Figure A4: Intensity of Russian interventions

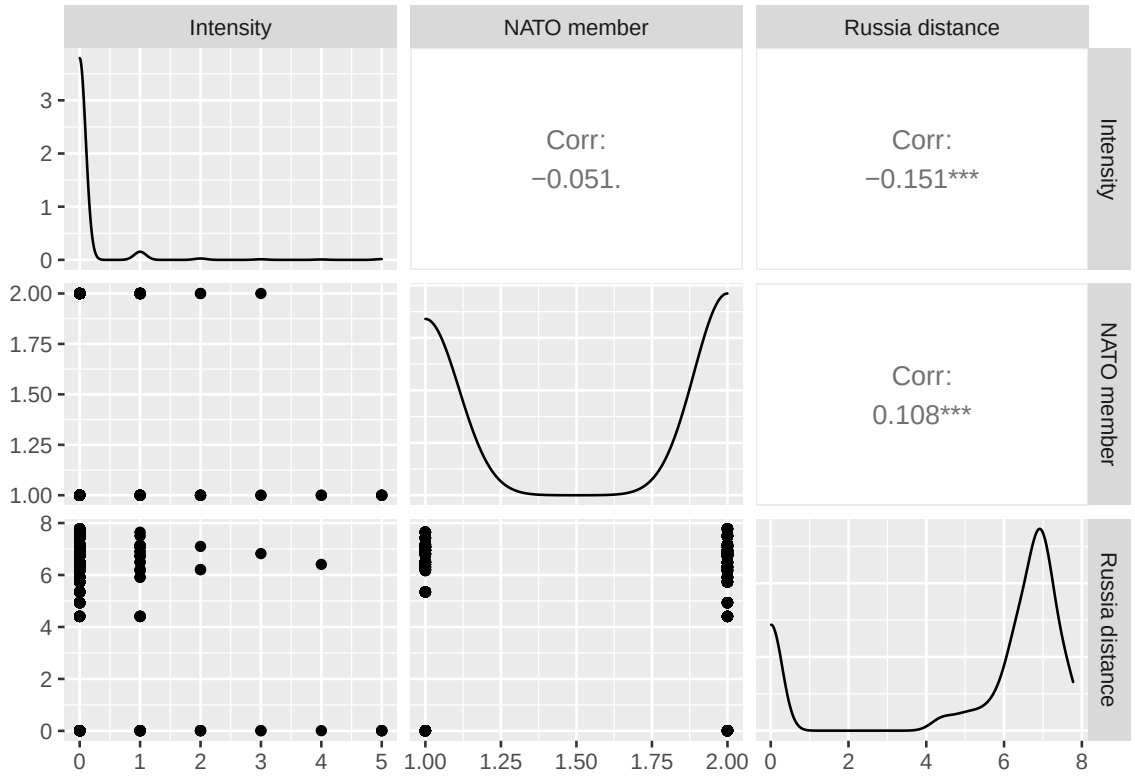

Figure A5: Bivariate correlation of dependent and independent variables

3.2 Odds ratios

Given the difficulty of interpreting ordered probit coefficients, Table 2 show the results as odds ratios with confidence intervals in parentheses when all other variables are held at their mean level. To use model 6 as an example for interpretation, for relevant NATO states, the odds of a non-cyber, non-information attack (categories 3, 4, or 5) versus a cyber attack, an information attack, or no attack is 58% lower.

	Full sample			Relevant states sample		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Independent Variables						
NATO member	0.76	0.63*	0.55*	0.63	0.56*	0.51*
	[0.39; 1.12]	[0.29; 0.96]	[0.19; 0.91]	[0.21; 1.05]	[0.14; 0.98]	[0.09; 0.93]
Russia distance	0.90*	0.90*	0.88*	0.95	0.92*	0.91*
	[0.84; 0.96]	[0.85; 0.95]	[0.83; 0.94]	[0.89; 1.01]	[0.86; 0.98]	[0.85; 0.98]
Controls						
Democracy		1.17	1.58		1.13	1.54
		[0.46; 1.88]	[0.89; 2.26]		[0.39; 1.86]	[0.84; 2.25]
Nuclear power		2.53*	1.56		2.50*	2.88*
		[1.85; 3.21]	[0.84; 2.28]		[1.72; 3.29]	[1.53; 4.23]
Population		1.21*	1.15		1.17*	1.20
		[1.06; 1.36]	[0.95; 1.35]		[1.01; 1.34]	[1.00; 1.41]
GDP per cap		0.99*	0.98*		0.99	0.99
		[0.98; 1.00]	[0.97; 1.00]		[0.98; 1.00]	[0.97; 1.00]
Mil. spending			1.02			1.00
			[1.00; 1.03]			[0.97; 1.02]
Observations	1,000	921	891	376	373	346

All models are ordered probits and include year-fixed effects with country-clustered standard errors in parentheses.

Table A3: Intensity of Russian Intervention: Odds Ratios

3.3 OLS regression

Although an ordered probit model is most appropriate given the dependent variable (intensity) is ordinal, we ensure that the sign on our coefficients are consistent with an OLS model that treats intensity as a continuous variable. See Table A4.

3.4 Ordered logit

We also run all models as ordered logits instead of ordered probits. Both are generalized linear models appropriate for an ordinal dependent variable that differ only in whether they use a logit link function as opposed to inverse normal link function (Johnston, McDonald, and Quist 2020). The results of the ordered logit in Table A5 are almost identical to those of the ordered probit, as expected.

3.5 Multiple imputation

Models 2, 3, 5, and 6 lose some observations due to missing values for control variables; primarily those not available after 2012. Variables with missing data are shown in Figure A6, with all but CINC being used in the models in the main text. We do not use the CINC ratio variable because listwise delection would lose 25% of our observations in a biased manner given the missingness is for all observations after 2012. Instead, the main model uses population and SIPRI military expenditure variables which adequately proxy for CINC given they are 2 of CINC's 6 components. When imputing missing values, we replace the population and military expenditure variables with CINC since it is more commonly used as an observable indicator for military power, the concept of interest.

	Full sample			Relevant states sample		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Independent Variables						
NATO member	-0.06^{***}	-0.07^{**}	-0.08^{***}	-0.14^{***}	-0.14^{***}	-0.16^{***}
	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)
Russia distance	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{**}	-0.02^{**}	-0.02^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Controls						
Democracy		−0.10	−0.06		−0.08	−0.02
		(0.12)	(0.13)		(0.13)	(0.14)
Nuclear power		0.12^{***}	0.09^{*}		0.14^{**}	0.11
		(0.04)	(0.05)		(0.06)	(0.10)
Population		-0.00^{**}	-0.00^{**}		-0.00^{**}	-0.00^{**}
		(0.00)	(0.00)		(0.00)	(0.00)
GDP per cap		0.02	0.01		0.04	0.04
		(0.01)	(0.02)		(0.03)	(0.04)
Mil. spending			0.00			0.00
			(0.00)			(0.00)
Observations	1,000	921	891	376	373	346

All models include year-fixed effects with country-clustered standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A4: OLS Results

	Full sample			Relevant states sample		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Independent Variables						
NATO member	−0.43	-0.97^{**}	-1.22^{**}	−0.74	-1.12^{**}	-1.29^{**}
	(0.50)	(0.46)	(0.49)	(0.54)	(0.53)	(0.54)
Russia distance	-0.18^{**}	-0.20^{***}	-0.21^{***}	−0.09	-0.17^{**}	-0.15^{*}
	(0.08)	(0.07)	(0.07)	(0.09)	(0.08)	(0.09)
Controls						
Democracy		0.79	1.27^{*}		0.64	1.13
		(0.82)	(0.76)		(0.82)	(0.74)
Nuclear power		1.70^{**}	0.97		1.67	2.41
		(0.82)	(0.88)		(1.03)	(1.88)
Population		0.48^{**}	0.42		0.42^{*}	0.50
		(0.22)	(0.28)		(0.24)	(0.30)
GDP per cap		-0.02^{*}	-0.03^{*}		−0.02	−0.02
		(0.01)	(0.02)		(0.02)	(0.02)
Mil. spending			0.02			−0.02
			(0.03)			(0.04)
Observations	1,000	921	891	376	373	346

All models include year-fixed effects with country-clustered standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A5: Ordered Logit Results

Missing values are calculated using bootstrap re-sampling across 10 different imputations using predictive mean matching (Buuren et al. 2006; White, Royston, and Wood 2011). The imputation predictions account for the temporal nature of the data. The same ordinal probit for each model is run across all 10 imputations

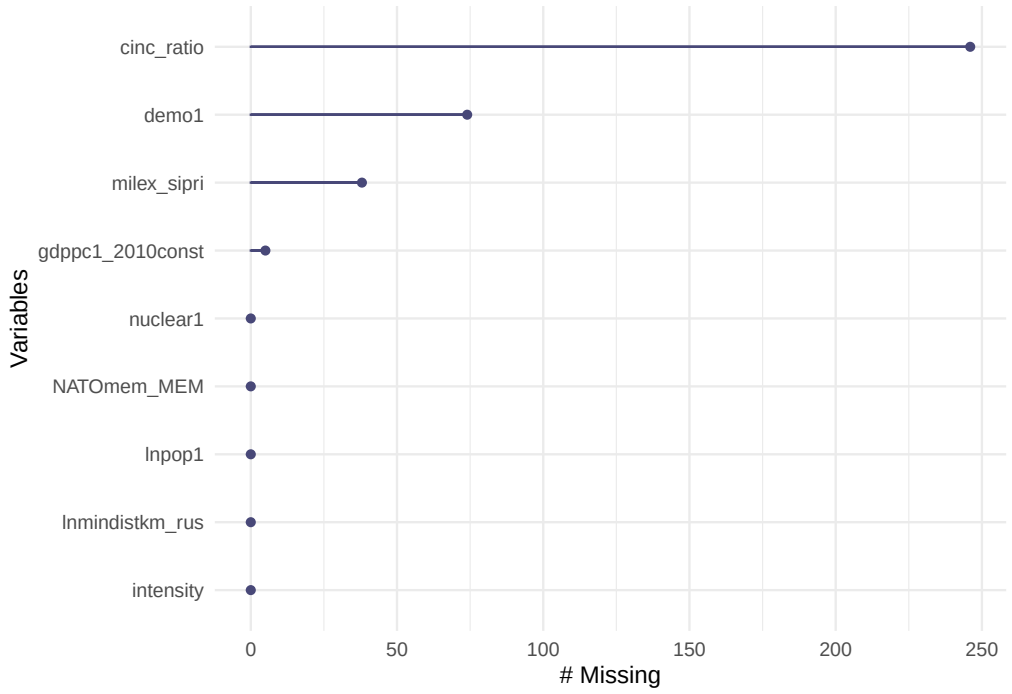

Figure A6: Number of missing observations for each variable in the dataset

and the coefficient estimates and standard errors are pooled across the 10 imputations to account for uncertainty produced by variation across the imputations. Variables not included in the main regression like ethno-linguistic fractionalization, and GDP per capita are included to increase the predictive performance of the imputation for variables that the literature suggests are correlated with the variables being imputed.

The results of the original models with imputed control variables are shown in Figure A7 and Table A6. We show the results for models 1 and 4 separately to make clear that these models have no imputed values. The coefficients in A6 are the same as those reported in models 1 and 4 in the main text and are only reported here to enable comparison with the imputed models in Figure A7.

	Full sample	Relevant states sample
	Model 1	Model 4
NATO member	0.76	0.63
	[0.39; 1.12]	[0.21; 1.05]
Russia distance	0.90*	0.95
	[0.84; 0.96]	[0.89; 1.01]
Observations	1,000	376

All models are ordered probits and include year-fixed effects with country-clustered standard errors in parentheses.

Table A6: Odds Ratios (non-imputed models)

3.6 Targeted states sample

We run the same models on a third sample of just targeted states as empirical support for the endogeneous bargaining and information asymmetry model extension offered in section 1.6 in the appendix. This includes only country-years that were targets of Russian aggression, meaning the intensity variable is greater than 0. Those results are shown in Table A7 and are consistent with the results produced by the other models.

Models with Imputed Control Variables

	Model 2	Model 3	Model 5	Model 6
Predictors	*Odds Ratios*	*Odds Ratios*	*Odds Ratios*	*Odds Ratios*
NATO member	0.39 **	0.37 **	0.28 **	0.29 **
	(0.16 – 0.94)	(0.16 – 0.87)	(0.09 – 0.85)	(0.11 – 0.79)
Russia distance	0.81 ***	0.81 ***	0.90 *	0.90 *
	(0.72 – 0.92)	(0.71 – 0.91)	(0.80 – 1.00)	(0.81 – 1.00)
Democracy	2.12	1.93	1.98	1.66
	(0.44 – 10.36)	(0.45 – 8.29)	(0.32 – 12.32)	(0.29 – 9.60)
Nuclear power	5.61 **	3.70 *	1.95	1.45
	(1.15 – 27.34)	(0.79 – 17.29)	(0.43 – 8.80)	(0.24 – 8.72)
GDP per cap	0.98	0.98 *	1.00	1.00
	(0.96 – 1.00)	(0.95 – 1.00)	(0.98 – 1.01)	(0.98 – 1.01)
Population	1.58 **		1.46 **	
	(1.04 – 2.42)		(1.02 – 2.10)	
CINC ratio		732.46 ***		120.07 *
		(12.60 – 42575.63)		(0.79 – 18277.02)
Observations	1000	1000	376	376
R^2	0.251	0.254	0.234	0.232

** p<0.1 ** p<0.05 *** p<0.01*

Figure A7: Intensity of Russian Intervention: Odds Ratios (Imputed models)

4 Case Study: US 2016

The main text presents case studies of Russian interventions in Estonia, Ukraine, and Georgia, which are all contiguous to Russia and former Soviet republics, and thus more comparable. Yet we expect the logic of the argument to apply more generally. Thus Russian intervention should be even more restrained against targets that are further away and more capable. Intervention in the 2016 U.S. election is consistent with this expectation.

A U.S. intelligence assessment released soon after the 2016 election concluded with "high confidence" that "Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the US presidential election. Russia's goals were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency. We further assess Putin and the Russian Government developed a clear preference for President-elect Trump" (Office of the Director of National Intelligence 2017). Moscow's influence operations might thus be described as unrestrained, even brazen, and thus motivated entirely by efficiency calculations. Yet the choice to pursue this course of action in the first place was very much constrained by the implicit deterrence posture of the United States. Russia could safely assume that the most powerful military in the world would retaliate for armed attacks against U.S. vital interests. While the United States had not designated its electoral process as "critical infrastructure" to explicitly signal that cyber interference was proscribed, Russia still had to consider the potential for American retaliation. Russia thus sought opportunities to impose costs and seek benefits while minimizing the risk of escalation. It found them through covert manipulation of democratic discourse. Indeed, Russia's electoral interference has gone essentially unpunished by the United States to date, aside from the expulsion of some Russian intelligence

	Model 1	Model 2	Model 3
Independent Variables			
NATO member	-2.16^{***}	-42.39^{***}	-43.40^{***}
	(0.64)	(1.86)	(1.72)
Russia distance	-0.19^{**}	-3.91^{***}	-2.19^{***}
	(0.08)	(0.14)	(0.08)
Controls			
Democracy		-83.35^{***}	-63.27^{***}
		(0.11)	(1.04)
Nuclear power		3.73	39.32^{***}
			(0.00)
Population		-19.79^{***}	-12.60^{***}
		(0.08)	(0.17)
GDP per cap		0.61^{***}	1.01^{***}
		(0.11)	(0.10)
Mil. spending			-1.67^{***}
			(0.36)

All models include year-fixed effects. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A7: Targeted States Sample

officers and the application of some additional sanctions to an already heavy regime put in place after Ukraine. Of course, if Trump's victory in 2016 or any of his administration's subsequent policies can ever be credited to active measures by the Russian Federation, even in part, it would amount to one of the most consequential intelligence coups in history. It is just as likely, however, that the Russian campaign simply added noise to one of the most chaotic campaigns in U.S. presidential history (Gelman and Azari 2017). Russian information operations appear to be a low-cost gamble to influence an over-determined outcome.

References

Abreu, Dilip, and Faruk Gul. 2000. "Bargaining and Reputation." *Econometrica* 68 (1): 85–117. https://doi.org/https://doi.org/10.1111/1468-0262.00094.

Acharya, Avidit, and Edoardo Grillo. 2015. "War with Crazy Types." *Political Science Research and Methods* 3 (2): 281–307. https://doi.org/10.1017/psrm.2014.23.

Brecher, Michael, and Jonathan Wilkenfeld. 1997. *A Study of Crisis*. University of Michigan Press.

Buuren, S. Van, J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, and D. B. Rubin. 2006. "Fully Conditional Specification in Multivariate Imputation." *Journal of Statistical Computation and Simulation* 76 (12): 1049–64. https://doi.org/10.1080/10629360600810434.

Casey, Adam, and Lucan Ahmad Way. 2017. "Russian Electoral Interventions, 1991-2017." Scholars Portal Dataverse. https://doi.org/10.5683/SP/BYRQQS.

Coe, Andrew J. 2011. "Costly Peace: A New Rationalist Explanation for War." Working Paper.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414. https://doi.org/10.1017/S0020818300033324.

Gartzke, Erik A. 1999. "War Is in the Error Term." *International Organization* 53 (3): 567–87. https://doi.org/10.1162/002081899550995.

Gelman, Andrew, and Julia Azari. 2017. "19 Things We Learned from the 2016 Election." *Statistics and Public Policy* 4 (1): 1–10. https://doi.org/10.1080/2330443X.2017.1356775.

Jackson, Matthew O., and Massimo Morelli. 2007. "Political Bias and War." *American Economic Review* 97 (4): 1353–73. https://doi.org/10.1257/aer.97.4.1353.

Johnston, Carla, James McDonald, and Kramer Quist. 2020. "A Generalized Ordered Probit Model." *Communications in Statistics - Theory and Methods* 49 (7): 1712–29. https://doi.org/10.1080/03610926.2019.1565780.

Office of the Director of National Intelligence. 2017. "Assessing Russian Activities and Intentions in Recent US Elections." Intelligence Community Assessment ICA 2017-01D. Washington, DC: National Intelligence Council. https://www.dni.gov/files/documents/ICA_2017_01.pdf.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60 (1): 169–203. https://doi.org/10.1017/S0020818306060061.

Valeriano, Brandon, and Ryan C Maness. 2014. "The Dynamics of Cyber Conflict Between Rival Antagonists, 2001–11." *Journal of Peace Research* 51 (3): 347–60. https://doi.org/10.1177/0022343313518940.

White, Ian R., Patrick Royston, and Angela M. Wood. 2011. "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." *Statistics in Medicine* 30 (4): 377–99. https://doi.org/https://doi.org/10.1002/sim.4067.