Appendix

Author names redacted

2020-11-13

Contents

This appendix provides supplemental information about the dataset of Russian gray zone campaigns introduced in the accompanying paper "After Deterrence: Explaining Conflict Short of War"

Formal Model

Formal statement of assumptions

First, we 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,* $\rho_0 < P(\tilde{g_C}, \tilde{g_D}) < 1$.[1]

Second, we express the assumption that if C's resolve increases, C becomes more willing to go to war over using gray zone conflict. As some intuition, conditional on gray zone conflict occurring, C selects one of two values for r. For the first value, the selected r will be the largest possible r that is tailored to keep D from going to war. I call this $\hat{g_C}$. For the second value, the selected g_C will be based on C's own resolve and represents the solution to C's internal optimization problem or C's internal efficiency. I call this $\check{g_C}$.[2] For

[1] 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 $\frac{1}{\beta_D} - \frac{\theta}{\beta_C} - 2\rho_0 + 2 > 0$ 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 $\frac{1}{\beta_D} - 4(\kappa_D - 1 + \rho_W) > 0$ if $\rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D} \leq \frac{\theta}{2\beta_C}$.

[2] Intuitively, $\tilde{g_C}$ is defined by $\tilde{g_C} = min\{\hat{g_C}, \check{g_C}\}$.

C's utility from war to be increasing in θ at a faster rate than the utility from gray zone conflict, we must consider both values of g_C.

Assumption 2: *The following must hold:* $\frac{d}{d\theta}\left[\theta\rho_W - \kappa_D - \left(\theta P(\hat{g_C}, \tilde{g_D}) - \beta(\tilde{g_D})^2\right)\right] > 0$ *and* $\frac{d}{d\theta}\left[\theta\rho_W - \kappa_D - \left(\theta P(\check{g_C}, \tilde{g_D}) - \beta(\tilde{g_D})^2\right)\right] > 0$.[3]

Proving Proposition 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

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

Here I discuss the intuition of the equilibrium in the paper. Assume for now that C is optimally selecting a g_C^* such that the game ends in gray zone conflict (in other words assume that $w_R^* = 0$ 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

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

I 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 $U_D = 1 - \rho_0 - g_C^* + \frac{1}{4\beta_D}$.

I can then begin considering C's utility. There are two things 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

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

which yields

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

Before discussing the true behavior, I want to 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

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

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, I can say that if C optimally selects into gray zone conflict, C will select $g_C^* = \tilde{g_C}$, where

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

I've characterized what happens withing gray zone conflict. I 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. I can calculate C's decision within the two cases of gray zone conflict:

First, I 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 $g_C^* = \hat{g_C} = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. I 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 is assumed 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 is assumed by Assumption 2.

Next, I 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 $g_C^* = \check{g_C} = \frac{\theta}{2\beta_C}$. I 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 is implied by Assumption 2.

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.

Equilibrium Behavior

Proposition 1A and the text below contains a more complete discussion on 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* $w_R^* = 0$ *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* $w_R^* = 1$*, 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* $w_R^* = 0$ *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* $w_R^* = 0$ *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* $w_R^* = 1$*. 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* $w_R^* = 0$ *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).

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 $-\kappa_D$.[4] 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 in 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.).

Extension 1: Endogenous β_D

In the model in the paper, I 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, I 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 lest D's costs from their gray zone response will be undefined.[5] The key take away from this extension is that if β_D is endogenous (and its selection costless), 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. Based on Assumptions 1 and 2, 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.[6]

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

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

- *1.A. I 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.*

[4] 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.

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

[6] 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.

- *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. I define* $\check{\beta_D}$ *implicitly as* $\theta = \frac{\kappa_C}{\left(\rho_W - \rho_0 - \frac{\theta}{4\beta_C} + \frac{1}{2\check{\beta_D}}\right)}$*. So long that* $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\check{\beta_D}}$*, 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 is not bound by D's deterrent threat.*

As one example of how this one equilibrium plays out, I 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 $2\beta_C(\rho_W - \rho_0 + \kappa_D + \frac{1}{4\dot{\beta_D}})$ 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. Essentially here, D is in a bad situation. 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.

Extension 2: Probabilistic Escalation to War

A useful feature of the model above is that everything that occurs is deterministic. It is only if a state wants to go to war or wants to enter gray zone conflict does it actually happen. However, this may not perfectly represent reality. 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 R, C will be more willing to accept the status quo or go to war.

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

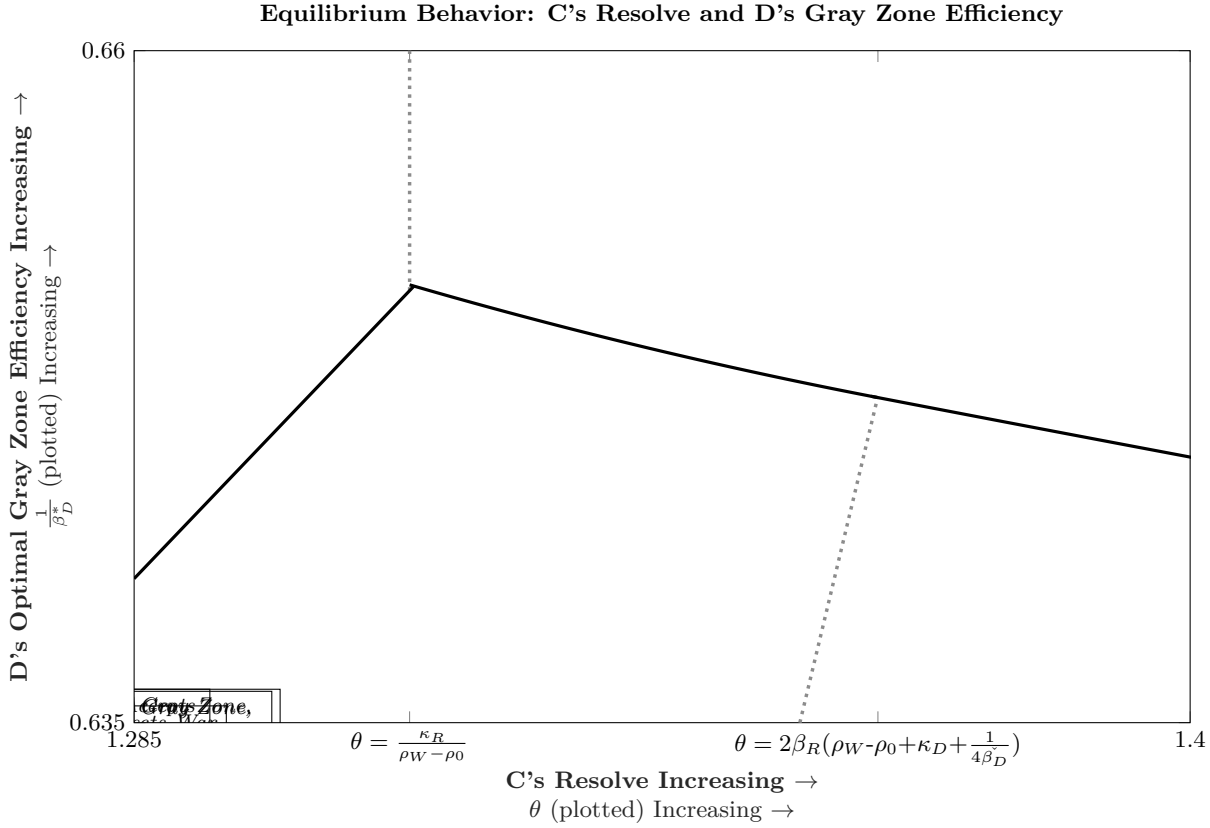


Figure 1: Extension 1: D's Optimal d^*

Figure 2: *

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 $\kappa_D = 0.1$.

$g_C > 0$, 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.

I express equilibrium behavior in Proposition B. Then below, I derive the new cut-points, Additionally in the derivations, I 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 $w_R^* = 0$ 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 $w_R^* = 1$.*
- *1.C. Otherwise, the game end in gray zone conflict where* C's *limited challenge is constrained by D's deterrent threat. C selects $w_R^* = 0$ 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 $w_R^* = 0$ 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 $w_R^* = 1$.*[7]
- *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 $w_R^* = 0$ and $g_C^* = \frac{\theta}{2\beta_C}$, and (assuming the game does not probabilistically escalate to war) D selects $w_D^* = 0$ and $g_D^* = \frac{1}{2\beta_D}$.*

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

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

[7] *While the right-hand-side of this condition is also increasing in θ, by Assumption 2, the left-hand-side increases faster with increases in θ.*

Note that the inequality sign does not flip because, by Assumption 1, $\rho_W - \rho_0 + \kappa_D - \frac{1}{4\beta_D} > 0$. I am 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 2 (as is written: that $\frac{1}{4\beta_D} - \kappa_D > 0$), the above sign does not flip. I 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, I 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 $g_C^* = \breve{g_C} = \frac{\theta}{2\beta_C}$. I 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 R.

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 $(1-\zeta)(\theta\rho_W-\theta\rho_0-\kappa_C)<0$. I can re-write the above as

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

Where 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 slip 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.

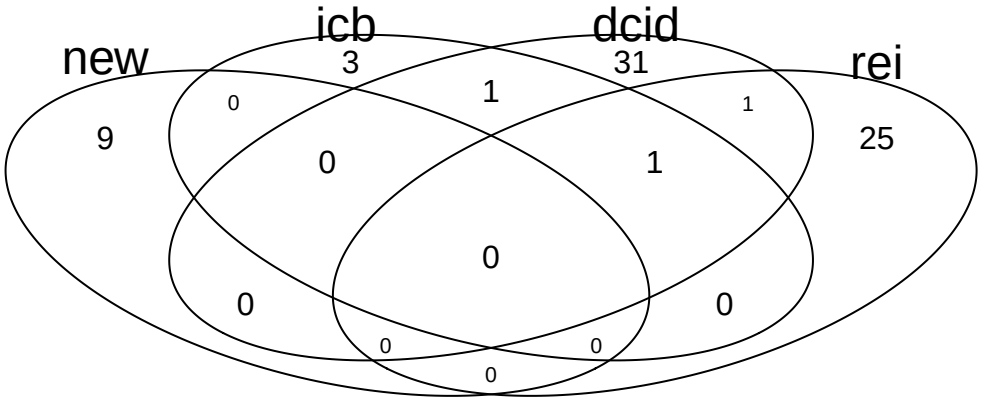
Finally, note that D's strategies in this game are unchanged from the game without probabilistic escalation.

New data

The universe of cases was created by first identifying cases of Russian foreign interventions from 3 prior datasets; ICB, DCID, and REI. 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.

Coverage of current datasets

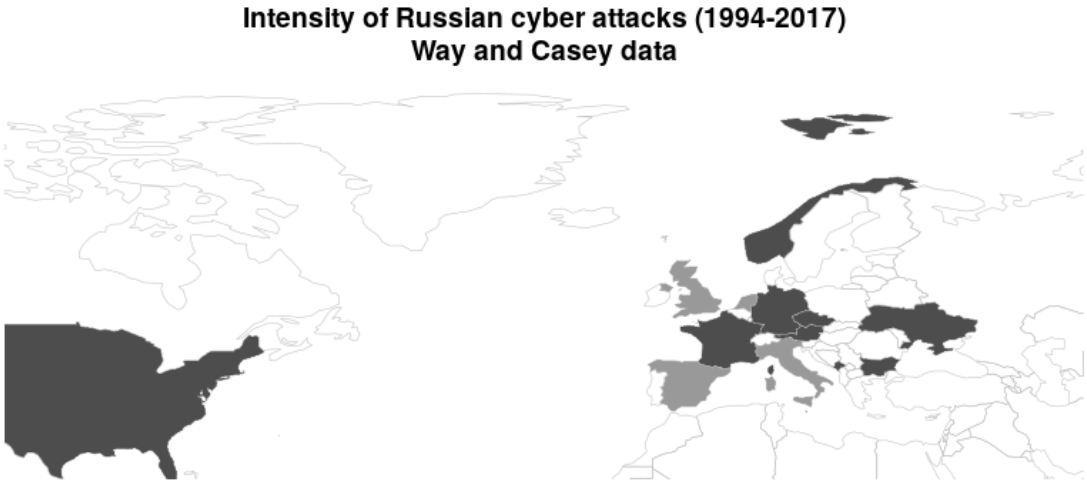
A comparison of what cases were covered in each individual dataset is provided here:



The overlap between cases is seen here:

Consistency of current 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) and REI (Way and Casey) is visualized here:



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.

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 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, etc
- **resp_cyberdisrup** - Did Russia use cyber attacks during this operation? That includes hacking, phishing, cyber espionage, DDOS attacks, etc
- **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.

Alternate model specifications