American Political Science Review

The Shadow of Deterrence: Why capable actors engage in conflict short of war

--Manuscript Draft--

Manuscript Number:	
Full Title:	The Shadow of Deterrence: Why capable actors engage in conflict short of war
Article Type:	Article
Section/Category:	International Relations
Keywords:	gray zone; deterrence; international security; international conflict; limited war
Corresponding Author:	J Andres Gannon UC San Diego: University of California San Diego RANCHO CUCAMONGA, California UNITED STATES
Corresponding Author's Institution:	UC San Diego: University of California San Diego
Corresponding Author's Secondary Institution:	
Order of Authors:	J Andrés Gannon, MA
	Erik A Gartzke, PhD
	Jon R Lindsay, PhD
	Peter Schram, PhD
Order of Authors Secondary Information:	
Abstract:	Recent conflicts have increasingly occurred in the ``gray zone'' between peace and open warfare. New technologies or tactics---from cyber operations to ``little green men''---may make aggression at low intensities more attractive to challengers (cheaper/more effective). Alternatively, existing deterrence networks may force motivated challengers to act more furtively (stability-instability). These dueling ``push-pull'' logics suggest contrasting conflict dynamics impacting stability and peace. We develop a game theoretic model to analyze gray zone conflict in which deterrence success is variable, rather than dichotomous. In the model, the scope and intensity of a challenger's provocation varies inversely with the implicit credibility of the defender's deterrent threat. We find empirical support for the stability-instability logic in Russian military actions since the 1990s; Russia is more restrained, and less effective, against nations in, or closely tied to, NATO. States face inherent trade-offs between stability and military potency in limiting the risk of escalation.
Suggested Reviewers:	
Opposed Reviewers:	
Additional Information:	
Question	**Response**
Please enter the total number of pages of your manuscript.	38
The following questions ask you about the ethics of the research you conducted for the manuscript you are submitting. **Unless otherwise noted, the information collected below will be shared with editors and reviewers, along with any appendix**	No

uploaded with the manuscript. Please anonymize the information as needed. In 2020, APSA Council approved *Principles and Guidance for Human Subjects Research*, which outlines several principles that should guide all political science research. The Principles include General Principles such as respect for the autonomy of participants (a wider category than "human subjects" as defined in US federal regulations), openness about ethical issues, the researcher's non-delegable responsibility for their ethical practices, and the reasoned justification of deviations from the Principles in scholarly publications, as well as additional specific Principles. The *APSR* editorial team adheres to these Principles. When submitting research for publication in the *APSR* that draws on direct engagement with human participants in the research process (see the Principles for clarification), you will be asked to affirm that your research conforms to all Principles and to state where you discuss the relevant aspects of your research practices (whether in the text or an appendix). If claiming an exception to one of the Principles, you will be asked to provide a reasoned justification for doing so. Did your research directly engage human participants in the research process?	
Did you and/or your co-authors encounter any ethical issues or perceptions of ethical issues unrelated to the study of human subjects?	No
Was this research funded by any agencies, organizations, or institutions?	Yes
Do you have any conflicts of interest with	No

the agencies, organizations, or institutions funding this research? as follow-up to "Was this research funded by any agencies, organizations, or institutions?"	
Transparency and openness considerations The following questions ask you about the transparency and openness of your research design. Unless otherwise noted, the information collected below will be shared with editors and reviewers. Please anonymize the information as needed. Did you generate or use qualitative or quantitative data for the research reported in this manuscript? Please select all that apply.	Yes, a quantitative dataset
Where do you describe how the *quantitative* data were generated or how you gained access to the data? Please check all that apply. as follow-up to "**Transparency and openness considerations** The following questions ask you about the transparency and openness of your research design. Unless otherwise noted, the information collected below will be shared with editors	In an appendix; In the main text

and reviewers. Please anonymize the information as needed. Did you generate or use qualitative or quantitative data for the research reported in this manuscript? Please select all that apply. "	
Do you agree to deposit in the *APSR* Dataverse at the conditional acceptance stage the quantitative datasets (and related documentation, interview guides, protocols, and scripts) and to have these materials published at the time the article or letter is published? as follow-up to "**Transparency and openness considerations** The following questions ask you about the transparency and openness of your research design. Unless otherwise noted, the information collected below will be shared with editors and reviewers. Please anonymize the information as needed. Did you generate or use qualitative or quantitative data for the research reported in this manuscript? Please select all that apply. "	Yes, without limitations
Author Comments:	We have provided here the minimum code and data needed to fully replicate the results of the paper. More comprehensive code for the entire project from start to finish is available in a github repository that will be made publicly accessible upon publication.
Manuscript Classifications:	International Relations; International Security

doi:xx.xxxx/xxxxx

The Shadow of Deterrence: Why capable actors engage in conflict short of war

J ANDRÉS GANNON *University of California, San Diego*
ERIK GARTZKE *University of California, San Diego*
JON R. LINDSAY *University of Toronto*
PETER SCHRAM *Vanderbilt University*

Recent conflicts have increasingly occurred in the "gray zone" between peace and open warfare. New technologies or tactics—from cyber operations to "little green men"—may make aggression at low intensities more attractive to challengers (cheaper/more effective). Alternatively, existing deterrence networks may force motivated challengers to act more furtively (stability-instability). These dueling "push-pull" logics suggest contrasting conflict dynamics impacting stability and peace. We develop a game theoretic model to analyze gray zone conflict in which deterrence success is variable, rather than dichotomous. In the model, the scope and intensity of a challenger's provocation varies inversely with the implicit credibility of the defender's deterrent threat. We find empirical support for the stability-instability logic in Russian military actions since the 1990s; Russia is more restrained, and less effective, against nations in, or closely tied to, NATO. States face inherent trade-offs between stability and military potency in limiting the risk of escalation.

Word Count: 11977

Corresponding Author: J Andrés Gannon (jagannon@ucsd.edu)

The authors wish to thank the members of the Center for Peace and Security Studies (cPASS), particularly Rex Douglass, as well as Nadiya Kostyuk, Chad Levinson, John Warden, and Christopher Whyte for thoughtful feedback. Tom Brailey, Cole Reynolds, Benjamin Smalley, and Erin Werner provided excellent research assistance. Earlier drafts of this paper were presented at the 2018 STRATCOM Deterrence Symposium and the 2018 American Political Science Association conference. This research was sponsored by Office of Naval Research Grant N00014-14-1-0071 and the Department of Defense Minerva Research Initiative. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. government. Data and code for the empirical analysis can be found at `https://github.com/CenterForPeaceAndSecurityStudies/GrayZone`.

This is a manuscript submitted for review.

doi:xx.xxxx/xxxxx

The Shadow of Deterrence: Why capable actors engage in conflict short of war

ANONYMISED AUTHOR(S) *Anonymised Institution(s)*

Recent conflicts have increasingly occurred in the "gray zone" between peace and open warfare. New technologies or tactics—from cyber operations to "little green men"—may make aggression at low intensities more attractive to challengers (cheaper/more effective). Alternatively, existing deterrence networks may force motivated challengers to act more furtively (stability-instability). These dueling "push-pull" logics suggest contrasting conflict dynamics impacting stability and peace. We develop a game theoretic model to analyze gray zone conflict in which deterrence success is variable, rather than dichotomous. In the model, the scope and intensity of a challenger's provocation varies inversely with the implicit credibility of the defender's deterrent threat. We find empirical support for the stability-instability logic in Russian military actions since the 1990s; Russia is more restrained, and less effective, against nations in, or closely tied to, NATO. States face inherent trade-offs between stability and military potency in limiting the risk of escalation.

Word Count: 11977

INTRODUCTION

In the wake of the downfall of Ukrainian President Viktor Yanukovych in February 2014, the Crimean Peninsula was invaded by "little green men," soldiers whose uniforms lacked insignia or other identifying information. The Kremlin formally annexed Crimea shortly thereafter. Protracted ground skirmishes, cyber campaigns, and "active measures" continue to plague Ukraine (Angevine et al. 2019). Russia's intervention has emerged as a paradigmatic example of a technologically novel and politically efficient form of "hybrid warfare," designed to challenge the status quo without triggering a broader conflict (Marten 2015; Lanoszka 2016; Chivvis 2017). Similar tactics and imagery have emerged elsewhere. For example, Chinese "little blue men" have been accused of eroding "red lines" in maritime East Asia (Green et al. 2017). According to former British Defense

Anonymised submission.

Secretary Michael Fallon (2017), "That is not a Cold War. It is a grey war. Permanently teetering on the edge of outright hostility. Persistently hovering around the threshold of what we would normally consider acts of war."

There is increasing concern that conventional conceptions of deterrence are inadequate to conceptualize and deal with burgeoning threats "in the gray zone" (Matisek 2017). Indeed, deterrence is typically believed to have failed if a challenger disrupts the status quo or resorts to military violence. As General Dunford (2016), Chairman of the United States Joint Chiefs of Staff, commented, "Our traditional approach is either we're at peace or at conflict. And I think that's insufficient to deal with the actors that actually seek to advance their interests while avoiding our strengths."

Rather than repudiating deterrence, however, gray zone strategies could actually reflect efforts by challengers to sustain existing frameworks, both to avoid major war and because they share key interests in common with status quo powers. The main source of discrepancy may lie in how we have come to think about deterrence, rather than how it is practiced. Deterrence helps to determine *how* a challenger disrupts, in addition to *whether*. Shaping other nations' behaviors could certainly prove valuable, even in the absence of peace and a full retention of the status quo. An enemy that engages "with one hand tied behind its back" to avoid triggering a larger contest is not fighting as effectively. Even if the challenger resorts to force and the defender does not intervene, fear of subsequent intervention by the defender could cause the challenger to adopt a more furtive (indecisive) military strategy.

We develop a formal model to differentiate and assess two distinct causal logics for gray zone conflict, one driven by military innovation (reduced costs or more effective aggression) and the other by deterrence. We also explore their distinctive empirical implications. In our model, a challenger and defender can select from variable intensities of limited conflict or choose to resolve the crisis with a decisive war. We show that the challenger's choice varies with the level of the deterrent threat posed by the defender, the challenger's valuation for the stakes, and the challenger's costs for fighting. We then test key results of the model, finding empirical evidence that the magnitude and risk of NATO intervention shapes recent Russian uses of force. We also find some (but weaker) evidence that Russia moderates the intensity of its efforts in response to the implicit credibility of the West's deterrence posture along an East-West gradient. These outcomes suggest Russian behavior is predicted by deterrence, and not only by the logic of gray zone as military innovation.

In the sections that follow, we first locate gray zone conflict in the broader literature on limited war (Section 2). We then analyze limited conflict using a formal model to illustrate the trade-offs that states face in deciding to enter into gray zone conflict or to go to war (Sections 3-5). Next, we assess our argument empirically, using data on Russian aggression, with an emphasis on cyber operations (Section 6). We then revisit the game theoretical results to discuss critical challenges of defense in the gray zone (Section 7). We conclude with implications of our argument (Section 8).

BETWEEN PEACE AND WAR

Despite the analytical convenience offered by conceptualizing war and peace as discrete outcomes, there is nothing new about conflict that falls ambiguously between peace and war (Lebow 2010). There is a long history of, and a vast literature on, limited war (Kissinger 1955; Osgood 1969), salami slicing tactics (Schelling 1966), low-intensity conflict (Turbiville 2002), revolutionary war (Shy and Collier 1986), military operations other than war (Kinross 2004), covert operations and proxy wars (Carson 2018; O'Rourke 2018), small wars (Olson 1990), hybrid wars (Lanoszka 2016), frozen conflict (Driscoll and Maliniak 2016), and hassling (Schram 2020).

Early Cold War writings on "weakening the enemy with pricks instead of blows" (Hart 1954, 186) emphasized limited political objectives in the shadow of nuclear escalation. The Korean War seemed a then-underappreciated type of war fought to achieve political ends short of total victory with military means short of a total commitment (Osgood 1969). Contemporary treatments understood limited war as a conflict between actors that had the capacity to increase battlefield commitments but did not want to do so, creating a third option between major war and acquiescence (Brodie 1957; Kissinger 1957). Strategists introduced the "stability-instability paradox" to describe how disincentives for nuclear war, or even major conventional war, encourage conflict at lower levels of intensity, or in peripheral theaters (Jervis 1984; Sagan and Waltz 2003). There is some threshold above which any given threat becomes too costly to be both credible and effective, and non-credibility invites challenges. As Snyder (1965, 167) observes, "nuclear technology introduced a new form of intent-perception and a new form of uncertainty — that concerning what types of military capability the opponent was likely to use and what degree of violence he was willing to risk or accept." Similarly, Powell (2015, 598) notes that "the amount of power the challenger brings to bear affects the stability of the conflict. More specifically, how much power the challenger brings to bear limits how much risk the defender can generate." As George and Smoke (1989, 173) explain, adversaries can "design around" deterrence by discovering new options that offer "an opportunity for gain while minimizing the risk of an unwanted response by the defender". This can result in serious fighting, as when Egypt "designed around" Israel's deterrent in 1973 (Stein 1989). Even so, "designing around" remains a perverse symptom of deterrence success because the adversary chooses its challenge in an effort to avoid the anticipated retaliation of the defender (Lieberman 2012).

Yet, other wars are limited not by the risk of escalation, but rather by cost concerns. During the Cold War there were numerous decolonization struggles and proxy wars in the developing world. In these "low intensity conflicts" or "small wars," the immediate adversaries tended to be irregular or guerrilla forces rather than peer competitors (Galula 1964; Taber 1965; Schultz 1986). An insurgent might give all, but still not be able to give much. Guerrillas with rudimentary arsenals and limited personnel simply could not directly engage powerful security forces, and thus opted for indirect ambush and subversion as a matter of necessity. After the Cold War, as great power competition waned and the United States became embroiled in occupations abroad, there was a revival of interest in questions of counterterrorism and counterinsurgency (Nagl 2005; Kilcullen 2010). Yet a common theme involves the limited military capacity of at least one of the combatants. Asymmetric contests thus contrast starkly with a superpower opting to forego military effectiveness to control escalation.

The renewal of interest in low-intensity conflict between more capable competitors represents a return to the earlier theme. A common thread in definitions of gray zone conflict is that it involves "a carefully planned campaign operating in the space between traditional diplomacy and overt military aggression" employed by challengers with grand geopolitical ambitions and potent capabilities (Mazarr 2015). A number of pundits highlight the growing diversity of technologies by which low intensity conflict can be practiced (Wirtz 2017). Even sceptics of cyber warfare highlight the expanded repertoire of means available for low-intensity conflict, especially online espionage and disinformation (Rid 2013; Jensen et al. 2019). If technological advances drive gray zone conflict, then we might expect to see

Russia engaging in it as often as possible.

However, militaries have been innovating novel technologies across multiple domains, at all levels of intensity. Thus the choice to use some of these innovations, but not others, really represents a restriction, rather than an expansion, of conflict behavior. Focusing on gray zone conflict as a reduction rather than expansion of options calls into question claims about its innovative effectiveness below the threshold of war. On the contrary, the familiar logic of the stability-instability paradox may be playing out today at different, and usually lower, thresholds (Lindsay and Gartzke 2018). In its classic formulation, nuclear deterrence paradoxically encourages peripheral conventional war. Today, the prospect of costly conventional war encourages provocation in cyberspace. The bad news about persistent conflict is thus good news about restraint.

In the last decade of the Cold War, US Secretary of State George Schultz (1986) offered a note of cautious optimism in this regard:

> The ironic fact is, these new and elusive challenges have proliferated, in part, because of our success in deterring nuclear and conventional war. Our adversaries know they cannot prevail against us in either type of war. So they have done the logical thing: they have turned to other methods. Low-intensity warfare is their answer to our conventional and nuclear strength a flanking maneuver, in military terms.

Below we develop Schultz's insight to explore whether modern gray zone conflict should be viewed as similarly reassuring or newly alarming.

THEORETICAL INTUITION

The Euromaidan Revolution presented Russia with a new political reality: Kiev was abruptly realigning with the West. Almost as quickly, Russia set about altering this new "status quo."[1] Russia's actions in Ukraine were limited and, because Russia stands to benefit from continued access to Crimea and from instability in Ukraine, politically advantageous. We refer to this behavior as "challenging the status

[1]The notion of "status quo" is inherently contextual, and thus subject to perspective and interpretation. Here, we treat the status quo as whatever political conditions prevail at a given point in time, regardless of past changes.

quo," or conducting "limited challenges" (implicitly to the status quo). In response to Russian activity, NATO increased its presence in the Black Sea, reinforced its support for capacity-building in Ukraine, and has stepped up its presence and cooperation in other countries in Eastern Europe. This has helped Ukraine in countering the Russian-sponsored separatist movement in the East of the country, and in dampening other related disruptions, such as cyber attacks.

The model we present in the next section captures these strategic dynamics. In the model, two actors, a challenger and a defender, experience a crisis. As the first move, the challenger decides whether to accept the status quo and do nothing, to conduct limited challenges, or to escalate to war. If the challenger conducts a limited challenge, the defender has an opportunity to respond with acceptance, escalation to war, or countering the challenge to the status quo with the defender's own limited conflict.[2] When the challenger engages in limited activity and the defender responds with a limited response (rather than accepting or going to war), we will say that the states are engaged in a gray zone conflict. Thus, gray zone conflict occurs when militarily capable challengers intentionally limit the intensity and capacity with which they conduct military operations, and the defender engages but chooses not to escalate to a decisive war. Narrowing the scope of gray zone conflict to militarily capable actors distinguishes the phenomenon from wars limited by cost by clarifying that actors are making a choice to limit the intensity of their engagement, rather then being compelled to do so by limited capabilities. Importantly, gray zone conflict must be preferred by both sides in a contest; both actors have the capacity to escalate to a larger war, but they choose not to. Escalation must thus be mutually undesirable, meaning that gray zone conflict is an equilibrium (Carson 2016, 2018).

Our model highlights the trade-offs confronting actors when choosing varying degrees of force. An actor may forgo the most effective or decisive means for the job when these means are too costly, they are not committed enough, or when some other more appealing alternative exists. While war can accomplish an aggressor's goals, it could also be unnecessary and inefficient if partial victories or *faits accomplis* can be achieved at lower cost, or risk (Altman 2018). By considering these trade-offs, our model offers insight concerning two central questions related to gray zone conflict. First, why do states

[2]The challenge and countering-the-challenge moves here are similar to "hassling" or "containment" as defined in Schram (2020) and Joseph (2020) (respectively), but are broader. For example, hassling specifically degrades a challenger's future military capabilities, which "challenges" could do here.

engage in gray zone conflict and, once they do, what explains variation in the intensity with which they pursue gray zone conflict? Second, when a defender faces the prospect of limited challenges, when does restrained behavior within the gray zone serve to benefit the defender?

Why do states engage in gray zone conflict? Challengers that attempt to alter the status quo through limited challenges must decide how aggressive the activity should be. A more intensive challenge can yield greater political gains. But the challenger faces two possible constraints into selecting limited challenges: an *external deterrent constraint*, and an *internal efficiency constraint*. If the *external deterrent constraint* binds, the challenger is resolved, but will limit the level of force to attempt to avoid a greater war. Thus, when the deterrent constraint binds, the challenger is most reactive to the defender's willingness to go to war. Alternatively, if the *internal efficiency constraint* binds, the challenger selects relatively low levels of force based on its own internal cost-benefit calculations. With the internal constraint, the challenger is most sensitive to its own valuation of the stakes versus the costs of challenging the status quo.

The theoretical distinction here is subtle, but empirically consequential. When the challenger's internal efficiency constraint binds, the challenger is able to choose the scope and intensity of conflict that it believes is most efficient for accomplishing its objective. The challenger can focus on pursuing the optimal challenge without concern of provoking an escalation, effectively wasting less effort for more political effect. When the challenger's external deterrent constraint binds, by contrast, the challenger must scale back from its optimal low-level challenge in order to avoid triggering a larger contest with the defender. In the latter case, the defender's willingness to go to war constrains the challenger to select from a set of less effective gray-zone options. A key empirical implication is that the intensity of gray zone conflict limited by deterrence should vary inversely with the credibility of the defender's deterrent posture, where the defender is more willing to absorb the costs of war. We label this moderating effect the defender's "deterrence gradient," encouraging greater provocations in areas where the defender is less resolved but more limited efforts where it is more so.

When is a restrained ability to counter gray zone conflict opportunistic? Within the game, we will assume that the defender incurs "gray zone costs" to countering limited challenges. A defender with low gray zone costs is very effective at engaging threats in the gray zone; thus, the defender having lower gray zone costs makes gray zone conflict less efficient for the challenger, which can encourage the challenger to pursue options outside of gray zone conflict. This can produce mixed results for the defender state. On one hand, lower gray zone costs for the defender could be productive for the defender if, upon the challenger's low-level activity becoming ineffective, the challenger abandons the use of force altogether and accepts the status quo. On the other hand, lower gray zone costs for the defender could be counterproductive for the defender if, upon the challenger's low-level activity becoming ineffective, the challenger instead escalates to war. Ultimately, the challenger's response to changes in the defender's costs and ability to conduct gray zone conflict will be arbitrated by whether the challenger prefers war to the status quo (or vice-versa), which is a function of the resolve of the challenger.

The model integrates some features from the formal literature on endogenous power shifts and deterrence (Fearon 1997; Debs and Monteiro 2014; Gurantz and Hirsch 2017; Baliga et al. 2020), specifically, a back-and-forth between a challenger and a defender. While this paper makes some simplifying assumptions regarding private information and hidden actions, it extends existing research by considering two competing actors, each with a continuum of policy options outside of declaring war or accepting peace. Far from just a technical flourish, this is critical for our results concerning the efficacy of a restrained ability to counter gray zone conflict. The model is thus most similar to those where politicians have flexible responses rather than just war or peace (Schultz 2010; Powell 2015; McCormack and Pascoe 2017; Coe 2018; Smith and Spaniel 2019; Spaniel 2019; Joseph 2020; Schram 2020) and where there is a strategic back-and-forth.[3]

[3]McCormack and Pascoe (2017) and Schram (2020) consider sanctions and hassling, respectively, as low-level options but do not provide the rival with an equivalent counter of implementing their own sanctions or hassling back.

MODEL AND EQUILIBRIUM

Game Form

We consider two states, a challenger C and defender D, that are in a crisis over a divisible asset with a normalized value of 1.[4] This model formalizes the following dynamic: two states are presented with a policy at the onset—a "status quo" policy—and then they decide whether and how to escalate. As intuition, the Euromaidan protests and overthrow of Yanukovych presented Russia with a new political reality, which can be thought of as the status quo policy in this model. Alternatively, the status quo policy could be viewed as an offered policy that arose through a series of steps in an not-modeled bargaining protocol where a bargaining failure may have occurred.

State C moves first. C either goes to war by setting $w_C = 1$, or sets $w_C = 0$. If C goes to war, the game terminates. If C sets $w_C = 0$, C also selects $g_C \in \mathcal{G}_C = \mathbb{R}_{\geq 0}$, where $g_C = 0$ is accepting the status quo, and $g_C > 0$ is conducting some limited, costly military action that shifts the political status quo in favor of the challenger. Second, as long as C did not previously go to war, State 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 D selects a gray zone response $g_D > 0$, D is using its own costly military means to weaken the impact of C's limited challenge. After the challenger acts and the defender responds, the game terminates, and payoffs are realized. For convenience, payoffs are summarized in Table 1.

If C goes to war at the outset (setting $w_C = 1$), C and D receive expected payoffs of $U_C = \theta\rho_W - \kappa_C$ and $U_D = 1 - \rho_W - \kappa_D$, respectively. These payoffs are largely consistent with the treatment of war as a costly lottery (Fearon 1997). $\kappa_C > 0$ and $\kappa_D > 0$ are C's and D's costs from war. $\rho_W \in [0, 1]$ is C's likelihood of winning in a war. The $\theta > 0$ term represents C's "resolve," or how much C cares about the asset in dispute.

If C sets $w_C = 0$ and $g_C = 0$, and D sets $w_D = 0$ and $g_D = 0$, this is equivalent to both states

[4] We recognize that many treatments of modern limited conflict focus on military aid to local proxies from a powerful patron (Plana 2020). For analytical parsimony, we consider a target's allies as part of the targets capabilities, discounted by the level of commitment (or disunity) in an alliance (Sobek and Clare 2013).

accepting the status quo $\rho_0 \in [0,1]$, and C and D receive payoffs $U_C = \theta\rho_0$ and $U_D = 1 - \rho_0$. We assume that $\rho_0 < \rho_W$, which implies that C is potentially dissatisfied with the status quo, and for a great enough resolve (θ), or low enough costs of war (κ_C), C will choose to go to war.

Now consider all outcomes where war does not occur (including the status quo), which is when C sets $w_C = 0$ and $g_C \geq 0$, and D sets $w_D = 0$ and $g_D \geq 0$. Here C and D receive payoffs $U_C = \theta P(g_C, g_D) - \beta_C g_C^2$ and $U_D = 1 - P(g_C, g_D) - \beta_D g_D^2$. The function $P : \mathcal{G}_C \times \mathcal{G}_D \to [0,1]$ represents the political outcome of gray zone conflict. We assume a functional form $P(g_C, g_D) = max\{min\{1, \rho_0 + g_C - g_D\}, \rho_0\}$, implying that P is weakly increasing in g_C and $-g_D$, and that P falls between ρ_0 and 1 inclusive. When $g_C = 0$ and $g_D = 0$, actors receive their status quo payoffs. By challenging the status quo ($g_C > 0$), C is shifting the status quo in its favor, but this can be dampened by D's selection of a gray zone response (when $g_D > 0$). That the final political settlement $P(g_C, g_D)$ must be weakly greater than the status quo ρ_0 implies that D's response to C's challenge cannot push C's final share of the asset to a level below what C is expected to attain from the status quo.[5] How C internalizes the final political outcome will depend on C's resolve, hence C's utility function having the $\theta P(g_C, g_D)$ term. Gray zone conflict is also costly to both actors. C pays costs $-\beta_C g_C^2$ for challenging, where $\beta_C > 0$. D pays $-\beta_D g_D^2$ for its gray zone conflict response. When β_D (β_C) is high, then D's (C's) costs of gray zone conflict are greater.

Finally, when D declares war after C engages in limited challenges (formally when $w_C = 0$, $g_C \geq 0$, and $w_D = 1$), C and D receive payoffs $U_C = \theta\rho_W - \kappa_C - \beta_C g_C^2$ and $U_D = 1 - \rho_W - \kappa_D$. Note that here C pays the costs of war as well as the costs of the limited challenge. Practically speaking, what is conducted for the purposes of gray zone conflict is different than what is conducted in a war, which makes the effort undertaken during gray zone conflict produce additional costs. Formally, these results do not change, provided a non-zero portion of the costs of C's limited challenge carry through.

[5]Challengers often do worse in wars than the status quo. For example, Germany started World War II with more territory than what it had at the end of the war. But adverse outcomes are much less likely with a limited disputes. In part, this is a rationale for our assumption that gray zone contests are less costly than major war.

TABLE 1. Summarized payoffs for actors

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

This model is designed to simply illustrate the comparative statics of how gray zone conflict plays out, conditional on intuitive cost and benefit parameters. In the Appendix, we consider two extensions. First, we examine an extension where β_D is endogenous. The new model offers little insight beyond what is discussed in Section 7. Second, we include a model where challenger activity creates probabilistic escalation. While this modification makes gray zone conflict less desirable to the challenger (and generates new equilibrium conditions), the results are substantively identical.

Equilibrium Concepts and Assumptions

We limit our attention to pure strategy subgame perfect Nash equilibria.[6] We use asterisks to denote equilibrium behavior, for example: g_C^* and g_D^*.

We make two simplifying assumptions (each is expressed formally in the Appendix). First, we limit analysis below to scenarios where, conditional on gray zone conflict occurring, the optimal limited challenge and response are such that the constraints on P do not bind (i.e. the final realized P is strictly less than 1 and strictly greater than ρ_0). This assumption is useful because it eliminates the possibility that kinks in the P function drive our results, and it prevents excessive casework.

Second, we assume that if C's resolve increases, C becomes more willing to go to war over using gray zone conflict. This is not to say that if C has a high resolve then C must use war; rather, it implies that the analysis is limited to a range of parameters where, as θ increases, C's utility from war increases at a faster rate than C's utility from gray zone conflict. Because limited challenges—and gray zone

[6]The focus on pure strategies negates cases where one player is indifferent over two actions (for example, selecting into gray zone conflict or selecting into war) and mixes over these actions.

conflict generally—can be thought of as more limited than war, this assumption rules out illogical cases where, *ceteris paribus*, challengers with intermediate levels of resolve go to war and challengers with high levels of resolve choose to undertake gray zone conflict.

Equilibria

The challenger decides whether the game ends with peace, war, or gray zone conflict. While the conditions characterizing C's choice are technical (as shown in Proposition 1), C's decision can be expressed in terms of C's level of resolve over the issue, θ. When C has low resolve, C will accept the status quo. When C has high resolve, C will select into war. When C's resolve falls between the two, assuming gray zone conflict is cost-effective enough, C will conduct limited challenges.

There is nuance in how C conducts gray zone conflict. To illustrate this, consider a hypothetical setting where C can select any intensity of limited challenge, and C knows that D will only respond with gray zone conflict (and not escalate to war). In this hypothetical setting, C faces an internal optimization, where C will select an intensity of limited challenge based on its resolve over the issue and its costs for conducting limited challenges—essentially selecting a limited challenge level where marginal returns are equal to marginal cost. This limited challenge based on C's internal cost-benefit analysis can be said to be determined by the challenger's *internal efficiency* constraints. Of course, in order not to select a limited challenge that is too aggressive and would cause the defender to escalate to war, C's limited challenge is also bound by the defender's *external deterrent threat* constraint. When C's optimal limited challenge is less aggressive than one that would provoke D to escalate, we say that C's internal efficiency binds. Otherwise, C will select a limited challenge tailored to make D refrain from war, and the deterrent threat binds. Which constraint binds is arbitrated by a technical condition—if $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$, or not.

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

Case 1, D's Deterrent Threat Binds, $\frac{\theta}{2\beta_C} \geq \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$*:*

- *1.A. C accepts the status quo (* $w_C^* = 0$ *and* $g_C^* = 0$*) 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* $\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}$.

- *1.B. C initially declares war ($w_C^* = 1$) 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 $\theta > \frac{\kappa_C}{\rho_W - \rho_0}$.*
- *1.C. C selects into gray zone conflict and is constrained by D's deterrent threat ($w_C^* = 0$ and $g_C^* = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$) otherwise.*

Case 2, C's Internal Efficiency Binds, $\frac{\theta}{2\beta_C} < \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$:

- *2.A. C accepts the status quo ($w_C^* = 0$ and $g_C^* = 0$) if $\theta \leq \frac{2\beta_C}{\beta_D}$ and $\theta \leq \frac{\kappa_C}{\rho_W - \rho_0}$.*
- *2.B. C initially declares war ($w_C^* = 1$) if $\theta > \frac{\kappa_C}{\rho_W - \rho_0 - \frac{\theta}{4\beta_C} + \frac{1}{2\beta_D}}$ and $\theta > \frac{\kappa_C}{\rho_W - \rho_0}$.*
- *2.C. C selects into gray zone conflict and is not constrained by D's deterrent threat ($w_C^* = 0$ and $g_C^* = \frac{\theta}{2\beta_C}$) otherwise.*

Proof: See Appendix.

We discuss the formal implications from Proposition 1 in two parts. First, in Section 5, we discuss what drives variation in gray zone activity, focusing on the comparative statics from D's deterrent threat (κ_D), C's resolve (θ), and C's cost of limited challenges (β_C). Then, in Section 7, we discuss why defenders may not always want to be effective at countering a challenger's gray zone activity, focusing our analysis of comparative statics on D's gray zone efficiency (β_D).

WHAT DRIVES VARIATION IN GRAY ZONE ACTIVITY?

On the Defender's Deterrent Threat and Conflict Intensity

As the defender becomes more willing to go to war (κ_D decrease), the challenger selects a weakly less aggressive limited challenges (g_C^*) or avoids gray zone conflict altogether. For example, if NATO leaders are more willing to escalate to war over NATO's core states relative to its periphery states or non-members, then the model expects less aggressive gray zone action against core states. The defender's willingness to go to war thus creates an upper bound on the tolerated level of limited challenges. If the defender becomes more willing to go to war, the challenger must either scale back the intensity of their limited challenge in order to avoid war, or the challenger must forgo limited challenges and either accept the status quo or go to war. If the defender's deterrent threat becomes less credible, the challenger has more freedom to choose whatever amount of force is needed to get the job done as the challenger sees fit.

Observation 1: *If the defender becomes more willing to go to war (lower κ_D's), then the challenger selects weakly less intense limited challenges, or the challenger may no longer engage in limited challenges and instead accepts the status quo or goes to war.*

We provide a formal discussion of Observation 1 in the context of Proposition 1 in the Appendix. Figure 1 illustrates one parameterized example of Observation 1. On the x-axis, moving left-to-right, D's deterrent threat from war is decreasing. This is operationalized in D's costs of war κ_D increasing. On the y-axis, moving low-to-high, C's optimal challenge g_C^* is increasing. For the lowest costs of war—the region where the equilibrium is described in Case 1.A in Proposition 1—D is very willing to go to war, C is tightly constrained in what limited challenges C can implement without provoking D to go to war, and thus gray zone conflict is not particularly productive for C. Altogether, under these parameters, C will select into the status quo.[7]

Moving to the right—to Case 1.C. equilibrium space—because D is less willing to go to war, C can engage in enough limited challenges to make gray zone conflict worthwhile. Here the intensity of C's limited challenge is constrained by D's deterrent threat, so C can select more aggressive limited challenges as D's deterrent threat decreases.

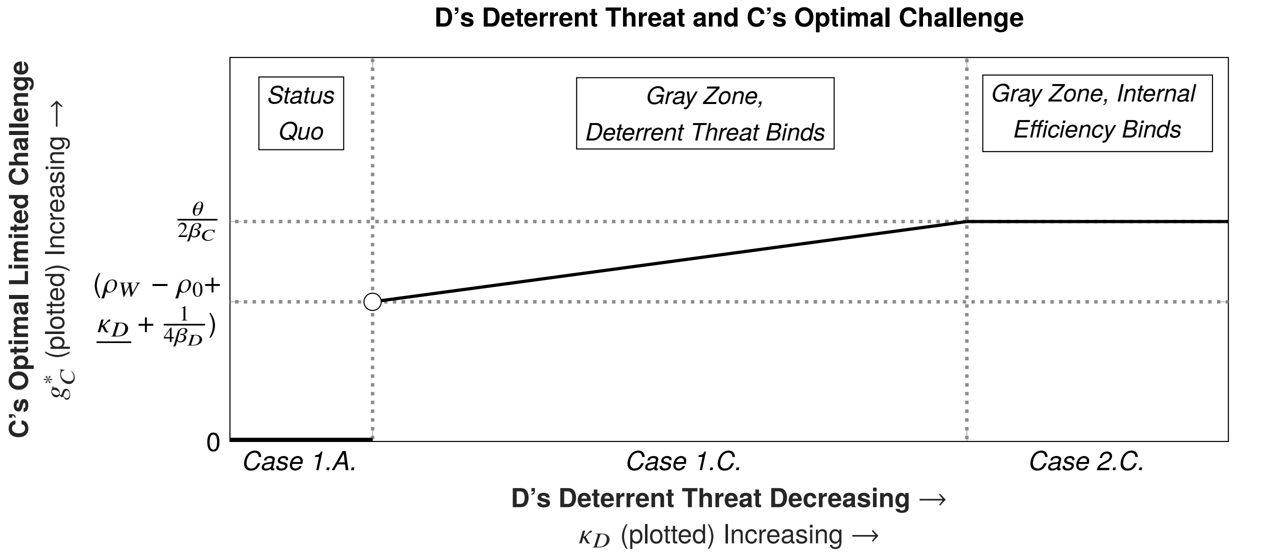
In Case 2.C, because D's deterrent threat has sufficiently declined, D is willing to tolerate aggressive limited challenges without escalating to war. Here, D's deterrent threat no longer binds, and instead C's limited challenge is constrained by C's internal efficiency constraint, which is not a function of D's deterrent threat.[8]

Altogether, Figure 1 illustrates a natural intuition. As D's deterrent threat from war decreases, C can select more aggressive limited challenges without provoking D to war. This expands the possible set of limited challenges that will not trigger an escalation, thus resulting in a greater intensity of gray zone conflict, until C's internal efficiency constraint binds and C no longer finds more gray zone conflict worthwhile.

[7]Note that if $\kappa_C = 0.41$, C would prefer going to war in this region rather than selecting the status quo. Also note that if we instead proxied D's deterrent threat using ρ_p, then C would not risk war in this region.

[8]The level of limited challenge that D would tolerate follows the trend of the diagonal line in the 1.C. region.

FIGURE 1. Optimal Limited Challenge as D's Deterrent Threat Decreases



Note: C's selected limited challenge intensity under a range of κ_D's are plotted. We define $\underline{\kappa_D}$ implicitly as the value of D's war costs that satisfies $\theta = \frac{\beta_C\left(\rho_W - \rho_0 + \underline{\kappa_D} + \frac{1}{4\beta_D}\right)^2}{\left(\rho_W - \rho_0 + \underline{\kappa_D} - \frac{1}{4\beta_D}\right)}$. All equilibrium cases are described in Proposition 1. The parameters are $\rho_0 = 0.1$, $\rho_W = 0.7$, $\beta_C = 0.34$, $\kappa_C = 0.5$, $\beta_D = 1$, and $\theta = 0.69$. To simplify labeling, y-axis not drawn to scale.

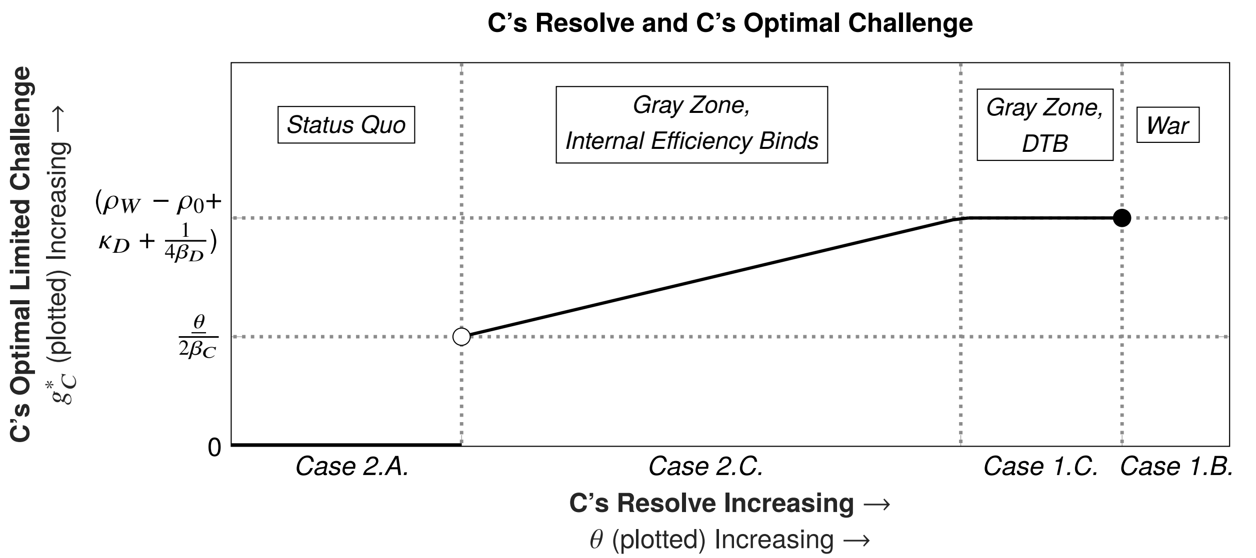
On the Challenger's Resolve and Conflict Intensity

As the challenger becomes more resolved or cares more about the policy issue (θ increases), the challenger's chosen level of limited challenge (g_C^*) weakly increases, unless its resolve is so high that it forgoes gray zone conflict and resorts to war. For example, during the Syrian Civil War, while both Moscow and the Syrian Ba'athist party clearly preferred retaining Bashar al-Assad's power, the Ba'athist party was presumably more resolved and fought harder than Russia. Consider a setting where the challenger optimally pursues a limited challenge to the status quo based on its own internal balancing of costs and benefits (the challenger's internal efficiency constraint binds). As the challenger's resolve increases, the challenger benefits more from shifting the status quo and is willing to select more aggressive limited challenges. However, a sufficiently highly resolved challenger could be willing to select a limited challenge intensity that exceeds the defender's tolerance within gray zone conflict. For these high levels of resolve, the challenger must either choose a non-internally-optimal intensity of its limited challenges to keep the defender from going to war—reflecting the external constraint of the defender's deterrent threat—or accept escalation to war.

Observation 2: *As the challenger's resolve increases, the challenger selects weakly more intense limited challenges, or may forgo gray zone conflict for war.*

Observation 2 follows naturally from Proposition 1, and Figure 2 illustrates one example of Observation 2. Moving from left-to-right, C's resolve θ is increasing. For the lowest resolve considered—the region where the equilibrium is described in Case 2.A.—C is not particularly resolved, and does not stand to gain sufficiently by altering the status quo. Because of this, C makes do with the status quo.

FIGURE 2. Optimal limited challenge as C's Resolve Increases



Note: C's selected limited challenge intensity under a range of θ's are plotted. We define $\underline{\theta} = \frac{2\beta_C}{\beta_D}$. "DTB" is an abbreviation for "Deterrent Threat Binds." All equilibrium cases are described in Proposition 1. The parameters are $\rho_0 = 0$, $\rho_W = 0.42$, $\beta_C = 1$, $\kappa_C = 0.5525$, $\beta_D = 1.53$, and $\kappa_D = 0.1$. To simplify labeling, y-axis not drawn to scale. We illustrate C selecting $w_A^* = 1$ as C selecting $g_C^* = 0$.

In Case 2.C, because C is more resolved to alter the status quo, C does best by engaging in limited challenges (gray zone). Here C's internal efficiency constraint binds, meaning C's selected limited challenge is increasing in C's resolve.

In Case 1.C, here C's resolve is even higher, but D's deterrent threat binds. C is resolved enough to engage in aggressive limited challenges beyond what D would tolerate (a level that would provoke D to escalate to war),[9] but C is not resolved enough to go to war. As a result, C selects a level of

[9]C's nominal willingness to challenge grows along the trend line from the Case 2.C.

limited challenge bound by D's indifference threshold between war and gray zone conflict, which is unchanging in C's resolve.

Finally, in Case 1.B, C's resolve has increased to the point where the level of limited challenges within gray zone conflict tolerated by D is not productive enough for C. Thus, in this region, C optimally goes to war over the issue.

On the Challenger's Gray Zone Costs and Conflict Intensity

As the challenger's costs from gray zone conflict decrease (β_C decreases), the challenger will select weakly more aggressive limited challenges (g_C^*). When the internal efficiency constraint binds, the challenger's selection of their limited challenges will increase as their costs decrease. When the external deterrent threat binds, the challenger's selection of their limited challenge does not vary with their costs, but rather is dictated by the defender's willingness to go to war, which is static.

Observation 3: *As C's gray zone costs decrease, C selects weakly more intense limited challenges, and may forgo accepting the status quo or war for gray zone conflict.*

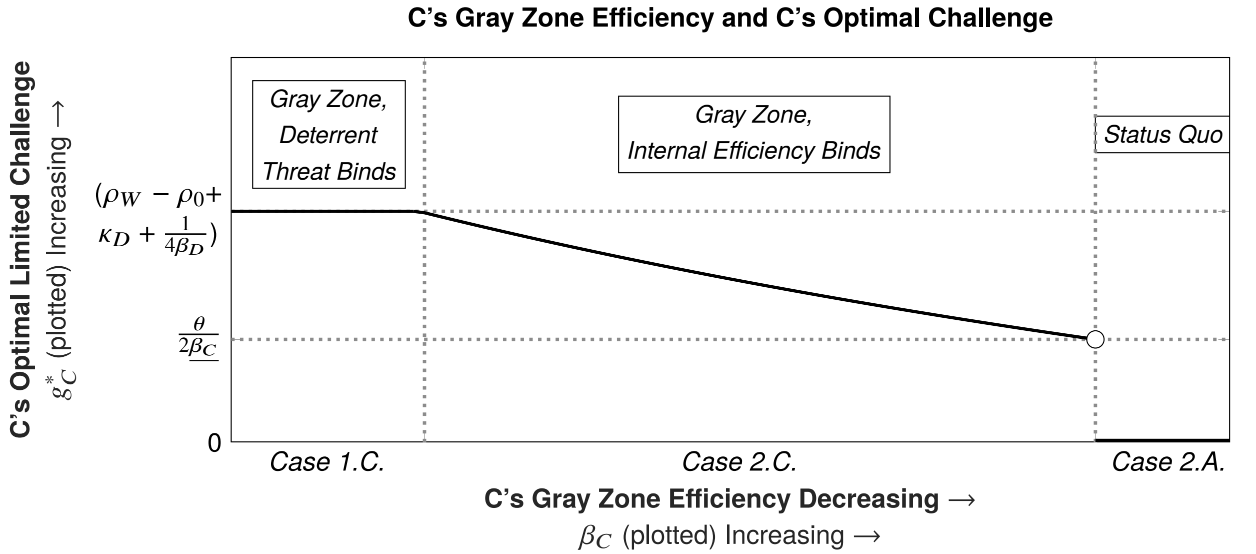
Figure 3 illustrates one example of Observation 3. Given the self-evident nature of this finding, we do not discuss these conditions at length.

EMPIRICAL APPLICATION: RUSSIAN EFFORTS IN THE GRAY ZONE

We empirically assess our argument by analyzing data on the scope and intensity of Russian foreign interventions over the past two decades. Quantitative analysis supports the hypothesis discussed in Observation 1: that Russia chooses its level of provocation in response to NATO's implicit deterrent threat in a given location. A secondary hypothesis allows us to operationalize the primary one: we expect the credibility of NATO deterrence to vary inversely with the military loss of strength gradient (Boulding 1962). Geography is not the focus of this article, but we use it as an instrument to map variation in the credibility of deterrence. We thus expect Russian military actions to appear more vigorous as the constraints of deterrence are relaxed.

We focus on Russia because its interventions are extensively referenced as paradigmatic examples

FIGURE 3. Optimal limited challenge as C's costs of limited challenges vary



Note: C's selected limited challenge intensity under a range of β_C's are plotted. We define $\underline{\beta_C}$ as $\frac{\theta}{2\underline{\beta_C}} = \rho_W - \rho_0 + \kappa_D + \frac{1}{4\beta_D}$. All equilibrium cases are described in Proposition 1. The parameters are are $\rho_0 = 0.1$, $\rho_W = 0.8$, $\kappa_D = 0.15$, $\kappa_C = 0.2$, $\beta_D = 1.25$, and $\theta = 0.2$. To simplify labeling, y-axis not drawn to scale..

of gray zone conflict (Marten 2015; Driscoll and Maliniak 2016; Chivvis 2017). From 1994 to 2018, Russia has been involved in election interference in the United Kingdom and Moldova, cyberattacks in Estonia and Georgia, and special operations in Ukraine and elsewhere. Russia's military adventures are the "most likely" cases for the argument that gray zone conflict is an effective and low-cost innovation for revising the status quo. The diversity of Russian targets and means employed provides an opportunity to conduct a controlled comparison of Russian choices under different deterrent circumstances. We present a simple, large-N statistical analysis where we test the implications of the model on Russian decision making while controlling for a number of relevant factors. By no means is this analysis causal, as our key independent variables—NATO membership, for example—are certainly not an exogenous treatment. Our results should be viewed as suggestive and should be taken with the necessary caveats. We then undertake a closer qualitative analysis of three cases: Estonia, Ukraine, and Georgia. We find that the possibility of a NATO intervention is associated with more limited gray zone operations, suggesting that NATO's deterrent threat shapes Russian gray zone behavior.

Data

Admittedly, data on Russian gray zone interventions are themselves a bit ambiguous. Most quantitative studies of gray zone operations focus on a particular type of gray zone activity, like cyber in the case of the Dyadic Cyber Incident and Dispute (DCID) data or electoral interference in the case of the Russian Electoral Interventions (REI) data (Valeriano and Maness 2014; Casey and Way 2017). Consequently, they cover almost entirely distinct samples with significant differences concerning the severity of Russian attacks.[10]

To address these deficiencies and discrepancies, we construct a new, expanded dataset including 82 cases of Russian intervention from 1994 to 2018. DCID and REI together describe 71 unique cases of Russian aggression that have either included some degree of cyber intervention or were cases of electoral interference. We identify 10 additional instances of Russian cyberattacks in the coding period that are not listed in previous datasets. Most of these new cases involve cyber conflict after 2011 (the latest year in DCID) that were non-electoral (the universe of cases in REI). We also include 3 cases of non-cyber Russian action from the International Crisis Behavior (ICB) dataset (Singer et al. 1972). For each incident, we create a new coding of the intensity of Russian attacks by coding whether Russia used five different types of military force in ascending order of intensity: (1) information operations (social media and disinformation), (2) cyber operations that result in disruption of infrastructure (service denial or industrial control system attacks), (3) overt use of special operations or unattributed military forces, (4) conventional air or sea forces, and (5) conventional ground forces.[11] This data is then aggregated to the country-year level with a coding for the highest level of intensity of Russian intervention against each country in each year. Our dependent variable is thus an ordinal variable coded 5 for the highest level of intensity down to 0 for country-years experiencing no Russian attack.

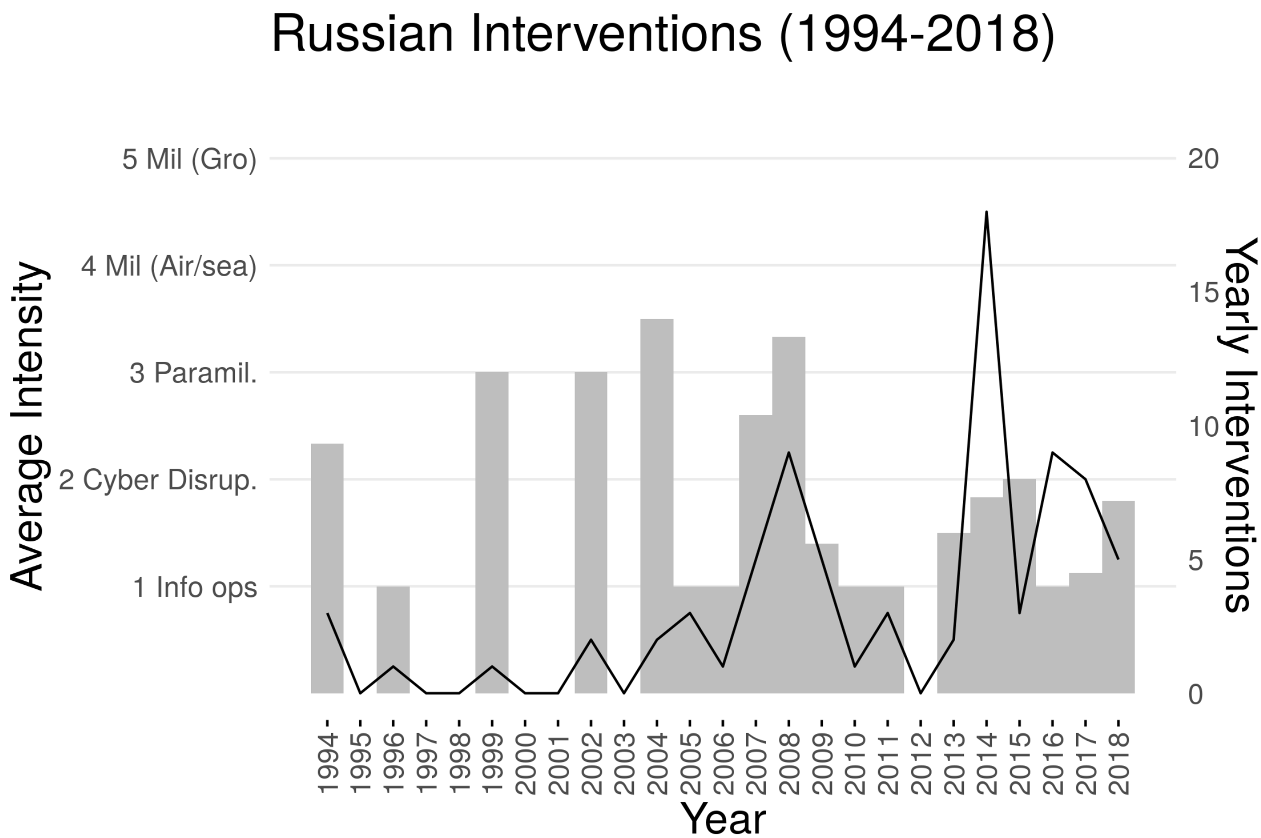
Figure 4 plots the count and average intensity of Russian gray zone operations since 1994. Contrary to descriptions of gray zone conflict as a new means of non-deterrable aggression or the product of an expansive technological portfolio, there does not appear to be a clear temporal pattern in the intensity

[10]Indeed, the only country-year that appears in both datasets is Ukraine 2014.

[11]We include conventional military interventions that occur alongside gray zone operations so as to avoid selecting on the dependent variable (Healey and Jervis 2020; Mareš and Netolická 2020).

or frequency of activity. Instead, 2004 represents the most intense overall Russian interventions and 2014 experienced the highest number of interventions (most of which were associated with Ukraine).

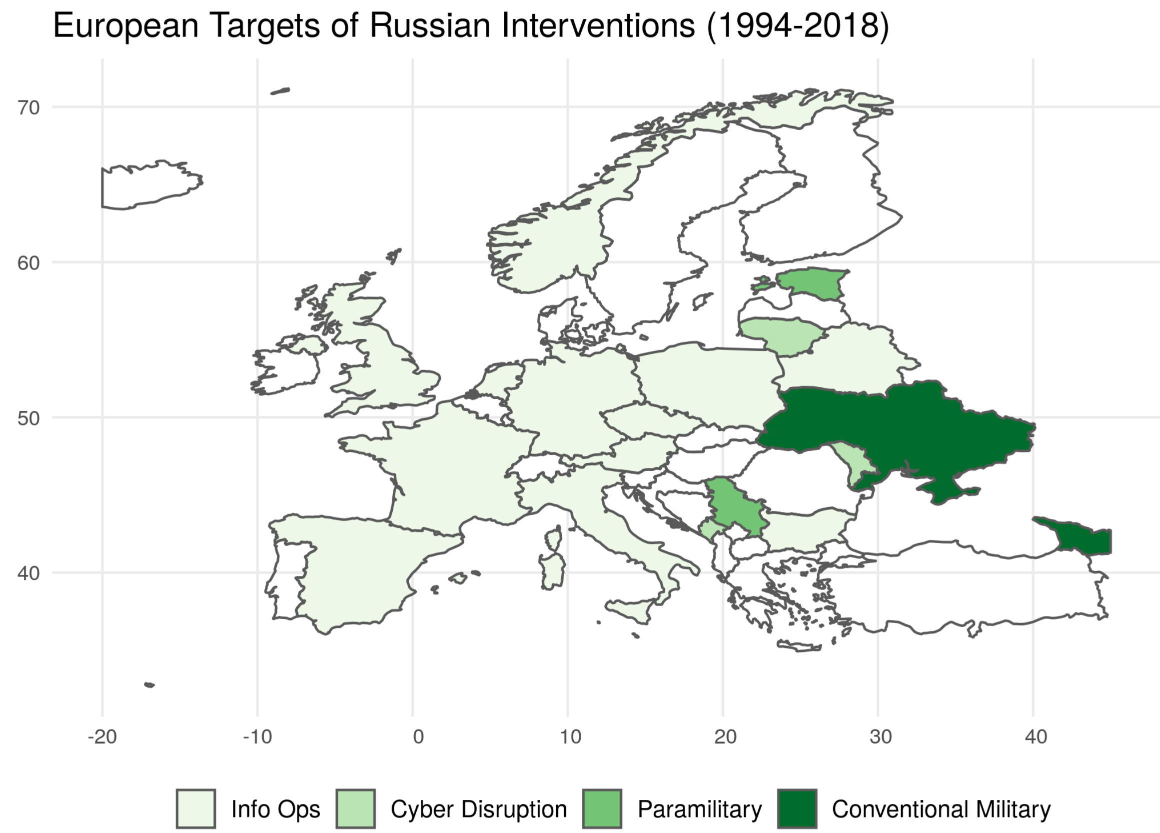
FIGURE 4. Intensity of Russian intervention across time



Note: The line represents the annual average of intensity. Bars denote the number of interventions per year.

Figure 5 depicts a pattern of the geographical coverage of Russian conflict events in Europe. Russia appears to be willing to use more force in countries in its "near abroad," relative to countries further away. Interpreting this geographic pattern on its own is difficult, as distance from Russia is plausibly related to Russian interest or resolve, Russia's ease of conducting operations, or the impact, or even the determinants of NATO membership. This pattern thus highlights the need for a more sophisticated statistical analysis.

FIGURE 5. Geographic representation of Russia intervention



Note: Shading represents the highest intensity of Russian intervention in each European state.

Now we discuss our independent variables. Consistent with the discussion on the external deterrent threat (Observation 1), we propose that the external deterrent threat from war is a key driver of Russian gray zone behavior. We operationalize this concept through a dummy variable for NATO membership. NATO members plausibly possess lower costs for fighting—since they can rely on collective security. They may thus have a reduced willingness to tolerate aggressive low-level behavior from Russia. If Russia is responding to NATO's deterrent threat, we expect NATO states to experience less intense Russian activity.

Consistent with the discussion of the internal-efficiency constraint (Observations 2 and 3), insofar as military power is affected by a loss of strength gradient (Corbett 1911; Posen 2003), we propose the intensity of Russian gray zone operations could decrease as Russia has less resolve over the issue, or faces greater costs for conducting operations.[12] We operationalize Russian resolve and gray zone efficacy jointly through a variable of the logged minimum distance between Russia and each potential-target state, as we expect Russia to care more about states on its periphery and to more easily conduct operations in its proximity.

We also include a series of control variables. We include a democracy dummy for states with a Polity V score greater or equal to 6 to control for potential Russian eagerness to target democracies (Early and Asal 2018). A state's possession of nuclear weapons may also alter Russia's calculus about how to pursue aggressive actions, so we include a dummy variable identifying whether a country-year is a nuclear state (Gartzke and Kroenig 2009). We include GDP per capita and the log of population as larger, richer states could afford more opportunities for Russian interventions, especially cyber interventions (Beckley 2010). Finally, we include military expenditure data from SIPRI since it influences both alliance decisions and the cost of undertaking aggression against an adversary (Omitoogun and Skons 2006). Summary statistics for all variables are included in the Online Appendix.

Model and Results

Because our outcome is an ordinal value, we estimate a series of ordered probit models with year fixed-effects and standard errors clustered by country. Our unit of analysis is the country-year. We run three empirical models on two samples. On the models, first we estimate the relationship between the intensity of Russian intervention and NATO membership and minimum distance from Russia without any control variables. Second, we re-run the first model while also controlling for the range of variables indicated above. We exclude military spending from the second model because it is missing across the entire panel for countries like Yugoslavia and Bosnia & Herzegovina. In model 3, we include the military spending variable. We operationalize military power using population and SIPRI data on military expenditure. On the samples, our first sample (models (1)-(3)) includes all European states. We define state membership using the Gleditsch and Ward state list and continent location using the World Bank Development Indicator (Gleditsch and Ward 1999). This excludes micro-states with less than

[12]We only analyze this variable in models that also include a NATO membership covariate.

250,000 people like Liechtenstein and San Marino. The downside is this sample may include states that are not of interest to Russia and may fall outside the scope of our model (for example, Luxembourg). To address this, our second sample (models (4)-(6)) represents "relevant European states," includes only European states that meet any of the three criteria: a) targets of a Russian attack from 1945-1993 as identified in the Militarized Interstate Dispute (MID) or International Crisis Behavior (ICB) datasets, b) Former Soviet Union or Warsaw Pact states, or c) states that are contiguous with Russia.

TABLE 2. Intensity of Russian Intervention: Ordered Probit Results

	Full sample			Relevant states sample		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Independent Variables						
NATO member	−0.28	-0.47^{**}	-0.60^{***}	-0.55^{**}	-0.68^{***}	-0.88^{***}
	(0.22)	(0.20)	(0.22)	(0.26)	(0.26)	(0.19)
Russia distance	-0.10^{***}	-0.11^{***}	-0.12^{***}	−0.03	-0.06^{**}	−0.03
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Controls						
Democracy		0.16	0.46		0.17	0.55
		(0.43)	(0.42)		(0.54)	(0.55)
Nuclear power		0.93^{**}	0.44		0.46	1.33^{*}
		(0.42)	(0.44)		(0.41)	(0.72)
Population		0.19^{**}	0.14		0.14^{*}	0.23^{***}
		(0.09)	(0.12)		(0.08)	(0.07)
GDP per cap		-0.01^{**}	-0.02^{**}		−0.00	−0.00
		(0.01)	(0.01)		(0.00)	(0.00)
Mil. spending			0.02			-0.03^{*}
			(0.01)			(0.01)
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 2 presents the coefficient estimates from the ordered probit regressions run on both samples. The results show that both NATO membership and distance from Russia decrease the intensity of Russian intervention against European states. Every models that either utilizes control variables (models (2), (3), (5), and (6)) or that samples on plausible relevant states (models (4)-(6)) suggests the relationship is statistically significant at least at the 0.05 level. Similarly, the coefficient for distance from Russia is in the expected direction in all models and statistically significant at the 0.05 level in models (1)-(3), and (5).

These results provide evidence of Observation 1: as the defender's deterrent threat increases, the challenger scales back the intensity of their limited challenges. For example, model (6) reports a proportional odds ratio of 0.42 on the NATO dummy.[13] This value means that 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. Our findings are also consistent with Observations 2 and 3, insofar as Russian valuation for the stakes and ease of operation arguably increases in regions deeper within its "Near Abroad" than for areas well beyond it. Together, this suggests that Russian behavior is shaped by the external deterrent threat in some cases and its own internal efficiency constraint in other cases, though the latter relationship is statistically weaker.

Our findings are consistent across a range of alternate samples and modeling specifications, as detailed in the Online Appendix. In additional models, missing values for control variables are replaced using multiple imputation with additive regression, bootstrapping, and predictive mean matching (van Buuren 2012). These results provide consistent statistical evidence in support of our argument, thus mitigating concern that the initial results are an artifact of listwise deletion (Lall 2017; Arel-Bundock and Pelc 2018). Additionally, we re-run the analysis including CINC ratios in place of population and the SIPRI military expenditure data (Singer et al. 1972).[14] To addressing missingness—CINC ratios are not published for years 2012-2018—we similarly use multiple imputation and find similar support for our argument.

If defender deterrence (and inversely challenger resolve) is conceived of as a continuous gradient, then the advent of low-cost means of aggression like cyber operations simply enables a challenger to fill in the lower regions of this gradient. At the same time, the systemic factors that lower the costs for cyber aggression also lower the costs of cyber countermeasures by the defender. We will return this idea later in the discussion.

[13]A complete table of all odds ratios is provided in the Online Appendix.

[14]Population and military expenditure are two of the six components that comprise the CINC index.

Case Studies

The quantitative analysis can be thought of as a coarse technique for considering empirical trends. Some concerns about endogeneity and causal inference can be partially addressed by examining the logic of Russian interventions in detailed qualitative case studies. For a more fine-grained test of our hypotheses, we consider three major cyber campaigns attributed to Russia that feature prominently in the cybersecurity literature: Estonia, Georgia, and Ukraine. We follow a most similar case study design in selecting cases that feature cyber attacks by the same contiguous challenger (Russia) but differ in other military instruments employed (Bennett and Elman 2007).[15] Additionally, to the extent that Russia wants to influence its immediate neighbors, Russia has an interest in all states and could intervene with relative ease. There are many potential explanations for why Russia wanted what it wanted in each instance, but here we set aside Russia's foreign policy formulation (Götz 2017; McFaul 2020). Instead, we highlight geo-strategic context and military effectiveness. A summary of the extent of Russian conflict behavior is provided in Table 3.

TABLE 3. Case comparison of Russian gray zone conflicts

Russian Response	Estonia (2007)	Ukraine (2014)	Georgia (2008)
Conventional Forces			X
Special Operations		X	X
Cyber Operations	X	X	X

Estonia (2007) Of the three states, Estonia experienced the most limited operations. Moscow coordinated a wave of DDoS attacks against Estonia following the relocation of a Soviet statue (Schmidt 2013). The gap in time between Estonia's 2004 ascension to NATO and the 2007 Russian cyber campaign is telling. In Georgia and Ukraine, the prospect of NATO ascension (announced in the April 2008 Bucharest Summit Declaration) provoked a Russian response. The Estonian attacks, by contrast, were muted and opportunistic, not a determined bid to change conditions on the ground. No one issued any clear demands or claimed responsibility, and Estonia did not replace the statue. The DDoS attacks were an ambiguous symbolic gesture calibrated to fall well below the threshold that might trigger a NATO response. The ambiguous legal status of a cyberattack in 2007 both enabled

[15] We include a fourth case study of Russian intervention in the 2016 US election in the appendix.

and constrained Russia in this respect (Joubert 2012). NATO was highly unlikely to escalate so long as Russia did not inflict serious harm. Estonia's defense minister considered but ultimately rejected invoking Article V, the collective defense clause of the NATO treaty, instead treating the episode as a domestic law enforcement matter (Traynor 2007). Overall, Russian moves seemed aimed at avoiding a greater escalation.

Ukraine (2014) Consistent with the logic of NATO's deterrent threat, Russian actions in Ukraine have been more extensive than those in Estonia, but less than what occurred in Georgia. Despite six years of protracted war there has occurred neither large-scale combined arms warfare, as in Georgia, nor unrestrained ethnic cleansing (Driscoll and Steinert-Threlkeld 2020). The fact that Russia could have exerted more effort, together with the actions made to allow both sides to save face, suggest Russian restraint.[16] Even though NATO has no formal commitment to Ukraine, conflict in a country that borders NATO allies like Poland and Hungary is implicitly shaped by the possibility of Western intervention, risking nuclear escalation in the process. As a result, we conjecture that Russia acts circumspectly. Endemic Russian cyberattacks and information operations have had little impact on battlefield events (Kostyuk and Zhukov 2019). Even as social media manipulation is supposedly a Russian specialty, pro-Kremlin narratives have not taken hold in Western Ukraine (Driscoll and Steinert-Threlkeld 2020). As Brantly et al. (2017) point out, despite notable diversity in the forms of conflict in Ukraine, they are neither intense nor effective enough to warrant outside intervention.

Georgia (2008) While Georgia was hit by DDoS service attacks (similar to Estonia) (Deibert et al. 2012), Russia also intervened militarily in South Ossetia and Abkhazia, an early example of cross-domain operations leveraging cyberspace. Russia's intervention choices in this conflict, situated at the far end of the Western deterrence gradient and deep in Russia's traditional sphere of influence, were relatively unconstrained. The same month that NATO announced a pathway to membership for Georgia, Russia announced that it would unilaterally increase peacekeepers in Abkhazia. Russia then used whatever mix of tools it needed to accomplish its objectives and did not pull its punches given that Western counteraction was unlikely (Binnendijk 2020). As Driscoll and Maliniak (2016, 590) point out,

[16] Mixed messages of resolve and restraint are common in covert action (Carson 2018).

"because of Georgia's location and its contested map, it is a security liability from the point of view of many in the West." The Russian intervention served to clarify the stakes of Western interference in its near abroad. While Russia's tactical performance left much to be desired, the mission was a strategic success that reinforced the status quo ante and ended the conversation about Georgia joining NATO. The forceful nature of the Russian intervention is notable, with long columns of conventional armor, something not considered elsewhere, despite the military imperatives of mass and firepower.

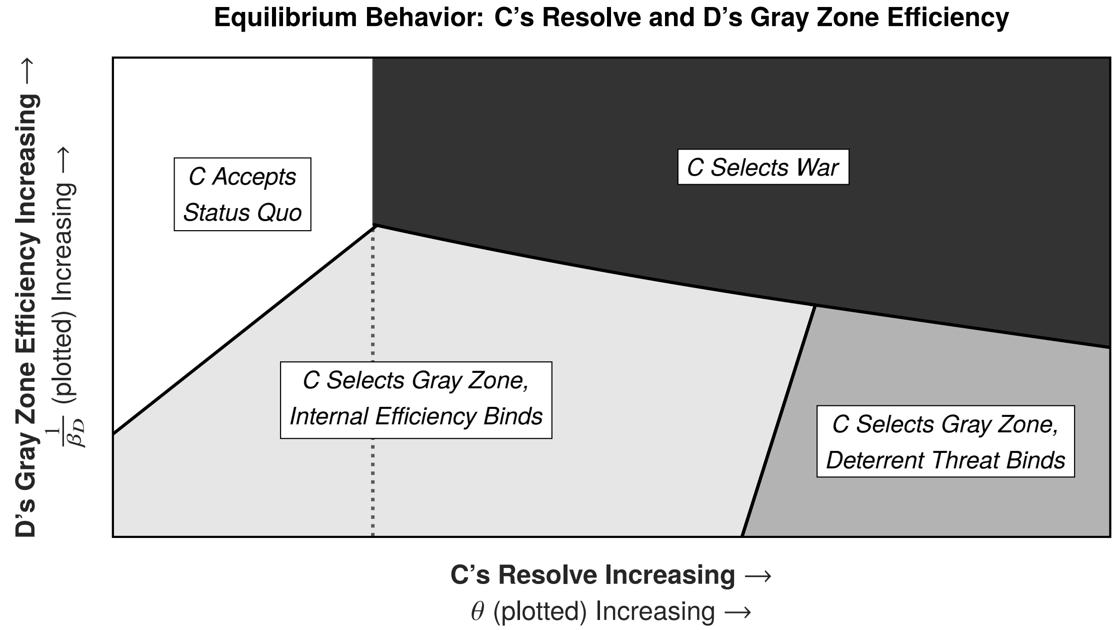
Discussion of Cases The overall pattern of recent Russian intervention is consistent with our hypothesis that deterrence encourages capable actors to engage in calculated restraint. Moving from Estonia to Ukraine and finally to Georgia, as the deterrent threat from NATO becomes less salient, Russia pursues its international objectives with greater intensity. One might argue that Russia has different levels of resolve across these cases. For example, one might argue that Russia places a very different value on the outcome in Ukraine than Estonia. Indeed, Russia let Estonia join NATO without a fight in 2004. By contrast, Russia had supported Georgian separatists since the early 1990s and was highly resolved to ward off Western encroachment. The Ukraine case, however, finds this alternative account wanting. The seat of the medieval Kievan Rus empire is arguably more salient in Russian nationalist mythology than Georgia, a peripheral outpost in the Caucuses far from Moscow, and the Black Sea port of Sevastopol also makes Crimea strategically important. If Russian moves were motivated by resolve rather than external deterrence, then we would expect more robust, and more overt, Russian military efforts in Ukraine. Yet, despite Russia's undoubted higher valuation for the stakes in Ukraine, one observes considerable restraint.

OTHER RESULTS: ESCALATION DYNAMICS

How can a defender best react to gray zone conflict? Is the answer to better prepare for such contests? Or, conversely, do improvements in the ability to counter gray zone conflict push a challenger into open warfare?

It might seem intuitive that improvements in defender capacity to counter gray zone aggression should always reinforce the strength of deterrence, but this is not necessarily the case. The final result from the model considers how changes in the defender's costs of gray zone conflict can alter strategic behavior, with some surprising implications for escalation. If the defender has low costs for conducting gray zone conflict, then it will be more effective at conducting gray zone operations, which leads the defender to behave more aggressively. If the challenger is thus comparatively worse in gray zone competition versus a capable defender, then the challenger may forgo gray zone conflict in favor either of the status quo or going to war. In other words, developing the tools to effectively thwart gray zone conflict could lead to greater escalation and more war.

FIGURE 6. Equilibrium behavior as C's resolve and D's gray zone efficiency varies



Note: C's resolve θ and the inverse D's gray zone efficiency $\frac{1}{\beta_D}$ are plotted. The dashed line is on the θ value where $\theta = \frac{\kappa_C}{\rho_W - \rho_0}$, or where C is indifferent between initially accepting the status quo and going to war. The parameters are $\rho_0 = 0$, $\rho_W = 0.5$, $\beta_C = 1$, $\kappa_C = 0.53$, and $\kappa_D = 0.1$.

Figure 6 details this logic. On the x-axis, we plot C's resolve. On the y-axis, we plot D's costs of gray zone conflict. Each region of the graph is an equilibrium type. For example, "C Selects Gray Zone, No Deterrent Threat" is described in Case 2.C. in Proposition 1. The dotted line represents the cut-point where, to the left, C prefers the status quo to war, and to the right, C prefers war to the status quo. Figure 6 conveys that D does not always benefit by increasing its ability to resist gray zone conflict. Consider a point in the "C Selects Gray Zone, Internal Efficiency Binds" region that is to the left of the dotted line. Here if D's gray zone efficiency increases, then after some point, C will forgo gray zone conflict altogether and accept the status quo. In this circumstance, C has a low-enough resolve that, with sufficient pressure, C can be deterred from all forms of conflict. However, consider now a point in the same equilibrium region, but to the right of the dotted line. If β_D increases above some point, C will choose war. In this circumstance, C is resolved enough that as D gets too good at countering gray zone conflict, C will opt into war.

Observation 4: *Decreases in D's gray zone costs (β_D) can lead to less or more conflict.*

There is an alternate interpretation; β_D could be influenced by the defender's prior, un-modeled moves in the game that set up current gray zone operations. For example, if the defender aggressively pursued counterinsurgency operations against foreign-backed rebels before this game began, the defender could be in a position to pursue aggressive gray zone conflict within the game. This interpretation illustrates how it is not just latent, exogenous costs that influence the challenger's activity, but rather war can resuly from the defender behaving too aggressively in prior actions against a highly resolved challenger.

This final interpretation has important policy consequences. Even if most actors are assumed to harbor challenger ambitions (Schweller 1996), would-be defenders still face a security-like dilemma in shaping how aggression is expressed. In the security dilemma, the outcome of a "threat" from the target is a function of whether or not a challenger is resolved. Conflict short of war complicates this picture because gray zone behavior by a highly resolved challenger and by a marginally-resolved challenger may be observationally indistinguishable. And yet, the consequence of the defender threatening to escalate within gray zone conflict vastly differs. The new U.S. Cyber Command doctrine of "persistent engagement" aims to establish dominance in strategic competition short of war through proactively "defending forward," but its very success could become the trigger for inadvertent escalation (Healey and Jervis 2020).

While we have focused on Russian gray zone conflict, we expect our theory to apply to conflict among capable powers more generally. Chinese incursions in the South China Sea offer another potential test. China's use of "little blue men" suggests that opportunism and restraint are both enabling and constraining Chinese foreign policy. That is, Beijing appears to fear that the use of more intense military operations risks provoking a Western response that both sides hope to avoid (Zhang 2019).

Focusing on the credibility of deterrence rather than the novelty of means used for gray-zone conflict can also help to evaluate proper policy responses (Green et al. 2017). Confronted with gray zone provocations by capable actors like Russia, China, and Iran, the United States would be well advised to reinforce its strengths while avoiding over-extension.

EVERY SILVER LINING'S GOT A TOUCH OF GRAY

Gray zone conflict occurs when capable actors intentionally limit the intensity or capacity of aggression and refrain from escalation. Deterrence shapes the way that conflict emerges, but it may not suppress conflict altogether. The good news is that gray zone conflict is symptomatic of deterrence success. The bad news is that gray zone conflict probes the threshold of deterrence effectiveness. We expect conflict severity to be greater wherever there are questions about the willingness or ability of defenders to respond forcefully. An adversary is seldom passive. There will always be attempts at end-runs or push-back, even when deterrence is credible. It is thus important to think carefully before overextending commitments where credibility is in doubt.

Just as there is a gray zone between war and peace, the distinction between effective and ineffective deterrence is also fuzzy. We have introduced the notion of a deterrence gradient, a straightforward extrapolation from the military loss of strength gradient, to describe credible deterrence as a continuous variable. Wherever deterrence is credible, or the challenger's resolve is low, challengers can be expected to exercise restraint as they probe to see what they can get away with. Wherever deterrence is not credible or challengers are highly resolved, however, a challenger must be more emboldened to use whatever means they have at their disposal to meet their objectives, limited only by internal efficiency constraints. The challenge lies between these extremes, where the variable threshold of credibility creates a policy arena for limited conflict, and where it can be difficult to distinguish efficiency motivations from risk sensitivity. Doubling down on deterrence can mitigate conflict in the latter case but provoke escalation in the former.

We have used the same cases that have raised alarms about the dangers of gray zone conflict—— Russian incursions in Georgia and Ukraine and cyber campaigns targeting many other countries——to suggest the validity of an alternative explanation. The evidence suggests that Russia systematically reduces the intensity of its interventions along the deterrence gradient, employing a greater variety of means with more lethal intensity where deterrence is weakest but conducting only ambiguous information operations where deterrence is most robust. The conventional wisdom is right that Russian interventions are paradigmatic exemplars of gray zone conflict, but it is wrong about Russian motivations and the effectiveness of these operations. Revisionist powers such as Russia have not discovered a secret formula or novel tools destined to destabilize Western democracies or undermine NATO's deterrence posture. Rather it acts opportunistically as circumstances enable it to hassle adversaries and their clients without, however, risking a costly military confrontation that Moscow does not want. The flip side of this logic, however, is that Russia is willing to call NATO's bluffs in cases where it can reasonably expect that NATO is unwilling to intervene. In Georgia, and even more so in Chechnya, Russian willingness to prioritize effectiveness at the price of efficiency is clear.

This argument has implications for the debate over NATO expansion after the Cold War (Shifrinson 2016). When expansion is posed in starkly binary terms, it can be seen as either a stabilizing force for Europe or an irresponsible provocation of legitimate Russian security interests fueled by liberal expansionism (McFaul et al. 2014; Mearsheimer 2014). If deterrence and conflict are continuous variables, however, then the real question is not simply whether NATO should or should not have expanded its security guarantees, but how far. One might thus argue that the first round of expansion to include the Eastern-Central countries (Poland, Hungary, Czech Republic) under the NATO umbrella helped to stabilize an historically conflict-prone portion of Europe. After the fall of the Soviet Union and during a period of military and economic weakness, moreover, Russia was grudgingly willing to accept a reduction in its European influence. One might also debate whether later rounds which brought in Baltic and Balkan countries made sense in whole or part. This is not the place to debate this history. We merely wish to point out that the alternative perspectives of NATO provocation and Russian aggression are better conceived of as context specific variables rather than absolute qualities of either actor. The right question is not whether NATO should have expanded, but how far.

Just as deterrence varies along the gradient, the contours of the gradient can shift over time. When NATO's relative power was increasing, expansion was defensible. If NATO's relative power decreases for whatever reason, then retrenchment makes more sense. Conversely, declining Russian relative power may enable NATO to bolster the line, rendering today's gray zone provocations more prohibitive tomorrow. Gray zone conflict allows keen observers to map the contours of the deterrence gradient, especially in areas where the "defender" has overreached its ability or will to respond. As information about the gradient is revealed, actors can take steps to shore up defenses and reassess priorities. Russia has advertised its willingness to interfere in elections, distort public debate, mobilize nationalist movements, and engage in other provocations. This in turn has led Western governments and publics to heighten awareness, increased vigilance, renewed defenses, and deterrence postures. Much as the shooting down of the Malaysian Airlines flight over Donetsk led both to renewed debate within NATO about intervention and to greater restraint in Moscow, so too the lowering of credible escalation thresholds can help to contain risk-averse opportunists. Just as gray zone conflict is symptomatic of deterrence success, the increasing incidence of Russian provocation may be symptomatic of a closing window for its effectiveness, such as it is.

The very fact that an adversary opts to engage in limited conflict suggests both vulnerabilities and opportunities. Instead of worrying that Russia is outwitting the West, we should instead realize that NATO has already blocked Russia from wielding even greater influence. The implicit general deterrence posture of NATO and the United States has arguably succeeded in keeping more extreme forms of Russian aggression in check. The unfortunate fact remains, however, that a simple remedy for gray zone conflict does not exist, because there is always some uncertainty about the precise contours of the deterrence gradient. We may be able to choose how our adversary confronts us, but not whether. Deterrence is not an on/off switch, but a rheostat, providing a range of causes and variable effects. Gray zone conflict is then not deterrence failure but a modest kind of success.

REFERENCES

Altman, Dan (2018). Advancing without Attacking: The Strategic Game around the Use of Force. *Security Studies 27*(1), 58–88.

Angevine, Robert , John K. Warden, Russell Keller, and Clark Frye (2019, May). Learning Lessons from the Ukraine Conflict. Technical Report NS D-10367, Institute for Defense Analyses.

Arel-Bundock, Vincent and Krzysztof J. Pelc (2018). When Can Multiple Imputation Improve Regression Estimates? *Political Analysis 26*(2), 240–245.

Baliga, Sandeep , Ethan Bueno De Mesquita, and Alexander Wolitzky (2020). Deterrence with Imperfect Attribution. *American Political Science Review 114*(4), 1155–1178.

Beckley, Michael (2010, February). Economic Development and Military Effectiveness. *Journal of Strategic Studies 33*(1), 43–79.

Bennett, Andrew and Colin Elman (2007, February). Case Study Methods in the International Relations Subfield. *Comparative Political Studies 40*(2), 170–195.

Binnendijk, Anika (2020, October). Understanding Russian Black Sea Power Dynamics Through National Security Gaming. Technical report, RAND Corporation, Santa Monica, CA.

Boulding, Kenneth (1962). *Conflict and Defense: A General Theory*. New York: Harper.

Brantly, Aaron , Nerea Cal, and Devlin Winkelstein (2017, December). Defending the Borderland: Ukrainian Military Experiences with IO, Cyber, and EW. Report, Army Cyber Institute.

Brodie, Bernard (1957). More About Limited War. *World Politics 10*(1), 112–122.

Carson, Austin (2016). Facing Off and Saving Face: Covert Intervention and Escalation Management in the Korean War. *International Organization 70*(1), 103–131.

Carson, Austin (2018). *Secret Wars: Covert Conflict in International Politics*. Princeton Studies in International History and Politics. Princeton, NJ: Princeton University Press.

Casey, Adam and Lucan Ahmad Way (2017). Russian Electoral Interventions, 1991-2017.

Chivvis, Christopher (2017, September). Hybrid war: Russian contemporary political warfare. *Bulletin of the Atomic Scientists 73*(5), 316–321.

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

Corbett, Julian (1911). *Some Principles of Maritime Strategy*. Longmans, Green and Co.

Debs, Alexandre and Nuno P. Monteiro (2014, January). Known Unknowns: Power Shifts, Uncertainty, and War. *International Organization 68*(1), 1–31.

Deibert, Ronald , Rafal Rohozinski, and Masashi Crete-Nishihata (2012, February). Cyclones in cyberspace: Information shaping and denial in the 2008 Russia–Georgia war. *Security Dialogue 43*(1), 3–24.

Driscoll, Jesse and Daniel Maliniak (2016). With Friends Like These: Brinkmanship and Chain-Ganging in Russia's Near Abroad. *Security Studies 25*(4), 585–607.

Driscoll, Jesse and Zachary Steinert-Threlkeld (2020). Social Media and Russian Territorial Irredentism: Some Facts and a Conjecture. *Post-Soviet Affairs 36*(2), 101–121.

Dunford, Joseph (2016, March). Gen. Dunford's Remarks and Q&A.

Early, Bryan and Victor Asal (2018). Nuclear weapons, existential threats, and the stability–instability paradox. *The Nonproliferation Review 25*(3-4), 223–247.

Fallon, Michael (2017, June). Speech delivered by Secretary of State for Defence Sir Michael Fallon at the RUSI Landwarfare Conference.

Fearon, James D. (1997, February). Signaling Foreign Policy Interests Tying Hands versus Sinking Costs. *Journal of Conflict Resolution 41*(1), 68–90.

Galula, David (1964). *Counterinsurgency Warfare: Theory and Practice*. Hailer Publishing.

Gartzke, Erik and Matthew Kroenig (2009, April). A Strategic Approach to Nuclear Proliferation. *Journal of Conflict Resolution 53*(2), 151–160.

George, Alexander and Richard Smoke (1989). Deterrence and Foreign Policy. *World Politics 41*(2), 170–182.

Gleditsch, Kristian S. and Michael D. Ward (1999, December). A revised list of independent states since the congress of Vienna. *International Interactions 25*(4), 393–413.

Götz, Elias (2017, June). Putin, the State, and War: The Causes of Russia's Near Abroad Assertion Revisited. *International Studies Review 19*(2), 228–253.

Green, Michael , Kathleen Hicks, Zack Cooper, John Schaus, and Jake Douglas (2017, May). *Countering Coercion in Maritime Asia: The Theory and Practice of Gray Zone Deterrence*. Rowman & Littlefield.

Gurantz, Ron and Alexander V. Hirsch (2017, May). Fear, Appeasement, and the Effectiveness of Deterrence. *The Journal of Politics 79*(3), 1041–1056.

Hart, Sir Basil Henry Liddell (1954). *Strategy: The Indirect Approach*. Faber & Faber.

Healey, Jason and Robert Jervis (2020). The Escalation Inversion and Other Oddities of Situational Cyber Stability. *Texas National Security Review*.

Jensen, Benjamin , Brandon Valeriano, and Ryan Maness (2019, January). Fancy bears and digital trolls: Cyber strategy with a Russian twist. *Journal of Strategic Studies 42*(2), 212–234.

Jervis, Robert (1984). *The Illogic of American Nuclear Strategy*. Cornell University Press.

Joseph, Michael F. (2020, March). A Little Bit of Cheap-Talk is a Dangerous Thing: States Can Communicate Intentions Persuasively and Raise the Risk of War in the Processes. *The Journal of Politics*, 709145.

Joubert, Vincent (2012). Five years after Estonia's cyber attacks: Lessons learned for NATO? Technical Report 76, NATO Defense College, Rome, Italy.

Kilcullen, David (2010). *Counterinsurgency*. Hurst.

Kinross, Stuart (2004, March). Clausewitz and Low-Intensity Conflict. *Journal of Strategic Studies 27*(1), 35–58.

Kissinger, Henry (1955). Military Policy and Defense of the "Grey Areas". *Foreign Affairs 33*(3), 416–428.

Kissinger, Henry (1957). Strategy and Organization. *Foreign Affairs 35*(3), 379–394.

Kostyuk, Nadiya and Yuri Zhukov (2019). Invisible Digital Front: Can Cyber Attacks Shape Battlefield Events? *Journal of Conflict Resolution 63*(2), 317–347.

Lall, Ranjit (2017). How Multiple Imputation Makes a Difference. *Political Analysis 24*(4), 414–433.

Lanoszka, Alexander (2016, January). Russian hybrid warfare and extended deterrence in eastern Europe. *International Affairs 92*(1), 175–195.

Lebow, Richard Ned (2010, September). The Past and Future of War. *International Relations 24*(3), 243–270.

Lieberman, Elli (2012, September). *Reconceptualizing Deterrence: Nudging Toward Rationality in Middle Eastern Rivalries*. Routledge.

Lindsay, Jon R. and Erik Gartzke (2018). Coercion through Cyberspace: The Stability-Instability Paradox Revisited. In K. M. Greenhill and P. Krause (Eds.), *Coercion: The Power to Hurt in International Politics*. New York, NY: Oxford University Press.

Mareš, Miroslav and Veronika Netolická (2020, August). Georgia 2008: Conflict Dynamics in the Cyber Domain. *Strategic Analysis 0*(0), 1–17.

Marten, Kimberly (2015, April). Putin's Choices: Explaining Russian Foreign Policy and Intervention in Ukraine. *The Washington Quarterly 38*(2), 189–204.

Matisek, Jahara W. (2017). Shades of Gray Deterrence: Issues of Fighting in the Gray Zone. *Journal of Strategic Security 10*(3), 1–26.

Mazarr, Michael (2015, February). Mastering the Gray Zone: Understanding a Changing Era of Conflict. Research Report, US Army War College, Strategic Studies Institute.

McCormack, Daniel and Henry Pascoe (2017, September). Sanctions and Preventive War. *Journal of Conflict Resolution 61*(8), 1711–1739.

McFaul, Michael (2020, October). Putin, Putinism, and the Domestic Determinants of Russian Foreign Policy. *International Security 45*(2), 95–139.

McFaul, Michael , Stephen Sestanovich, and John Mearsheimer (November/December 2014). Faulty Powers. *Foreign Affairs*.

Mearsheimer, John (September/October 2014). Why the Ukraine Crisis Is the West's Fault. *Foreign Affairs*.

Nagl, John (2005). *Learning to Eat Soup with a Knife: Counterinsurgency Lessons from Malaya and Vietnam*. University of Chicago Press.

Olson, William (1990, April). The concept of small wars. *Small Wars & Insurgencies 1*(1), 39–46.

Omitoogun, Wuyi and Elisabeth Skons (2006). Military expenditure data: A 40-year overview. In *SIPRI Yearbook 2006: Armaments, Disarmament and International Security*, pp. 269–294. Stockholm International Peace Research Institute.

O'Rourke, Lindsey (2018). *Covert Regime Change: America's Secret Cold War*. Cornell Studies in Security Affairs. Ithaca, NY: Cornell University Press.

Osgood, Robert (1969, February). The reappraisal of limited war. *The Adelphi Papers 9*(54), 41–54.

Plana, Sara (2020, March). Proxy War: The "Least Bad Option" or the "Second-to-Last Resort"? *Texas National Security Review*.

Posen, Barry (2003, July). Command of the Commons: The Military Foundation of U.S. Hegemony. *International Security 28*(1), 5–46.

Powell, Robert (2015, June). Nuclear Brinkmanship, Limited War, and Military Power. *International Organization 69*(3), 589–626.

Rid, Thomas (2013). Cyberwar and Peace. *Foreign Affairs*.

Sagan, Scott and Kenneth Waltz (2003). *The Spread of Nuclear Weapons: A Debate Renewed*. Norton.

Schelling, Thomas (1966). *Arms and Influence*. Yale University Press.

Schmidt, Andreas (2013). The Estonian Cyberattacks. In J. Healey (Ed.), *A Fierce Domain: Conflict in Cyberspace, 1986 to 2012*, pp. 174–193. Cyber Conflict Studies Association.

Schram, Peter (2020, June). Hassling: How States Prevent a Preventive War. *American Journal of Political Science*.

Schultz, George (1986, January). Low-Intensity Warfare: The Challenge of Ambiguity.

Schultz, Kenneth A (2010). The enforcement problem in coercive bargaining: Interstate conflict over rebel support in civil wars. *International Organization 64*(02), 281–312.

Schweller, Randall (1996, March). Neorealism's status-quo bias: What security dilemma? *Security Studies 5*(3), 90–121.

Shifrinson, Joshua R. Itzkowitz (2016, April). Deal or No Deal? The End of the Cold War and the U.S. Offer to Limit NATO Expansion. *International Security 40*(4), 7–44.

Shy, John and Thomas Collier (1986). Revolutionary War. In P. Paret, G. Craig, and F. Gilbert (Eds.), *Makers of Modern Strategy from Machiavelli to the Nuclear Age*, pp. 815–862. New Jersey: Princeton University Press.

Singer, David , Stuart Bremer, and John Stuckey (1972). *Capability Distribution, Uncertainty, and Major Power War, 1820-1965*, pp. 19–48. Sage Publications.

Smith, Bradley C. and William Spaniel (2019, May). Militarized Disputes, Uncertainty, and Leader Tenure. *Journal of Conflict Resolution 63*(5), 1222–1252.

Snyder, Glenn (1965). The Balance of Power and the Balance of Terror. In F. Hartmann (Ed.), *World in Crisis: Readings in International Relations*, pp. 180–191. New York: The Macmillan Company.

Sobek, David and Joe Clare (2013, July). Me, myself, and allies: Understanding the external sources of power. *Journal of Peace Research 50*(4), 469–478.

Spaniel, William (2019). *Bargaining over the bomb: the successes and failures of nuclear negotiations*. Cambridge University Press.

Stein, Janice Gross (1989, April). *Calculation, Miscalculation, and Conventional Deterrence*. JHU Press.

Taber, Robert (1965). *War of the Flea: The Classic Study of Guerrilla Warfare*. L. Stewart.

Traynor, Ian (2007, May). Russia accused of unleashing cyberwar to disable Estonia. *The Guardian*.

Turbiville, Graham (2002, June). Preface: Future Trends in Low Intensity Conflict. *Low Intensity Conflict & Law Enforcement 11*(2-3), 155–163.

Valeriano, Brandon and Ryan C Maness (2014, May). The dynamics of cyber conflict between rival antagonists, 2001–11. *Journal of Peace Research 51*(3), 347–360.

van Buuren, Stef (2012, March). *Flexible Imputation of Missing Data*. CRC Press.

Wirtz, James J. (2017, April). Life in the "Gray Zone": Observations for contemporary strategists. *Defense & Security Analysis 33*(2), 106–114.

Zhang, Ketian (2019, July). Cautious Bully: Reputation, Resolve, and Beijing's Use of Coercion in the South China Sea. *International Security 44*(1), 117–159.

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

Author names redacted

2020-12-07

Contents

This appendix accompanies the paper "The Shadow of Deterrence: Why capable actors engage in conflict 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

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. This is $\hat{g_C}$. 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. This is $\check{g_C}$.[2] For 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]

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

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

Here we 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 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

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

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

[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 $-\kappa_D + \frac{1}{4\beta_D} > 0$.

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, 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 withing 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 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, 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 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.

1.2.2 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).

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 $-\kappa_D$.[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.).

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 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

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

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

[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.

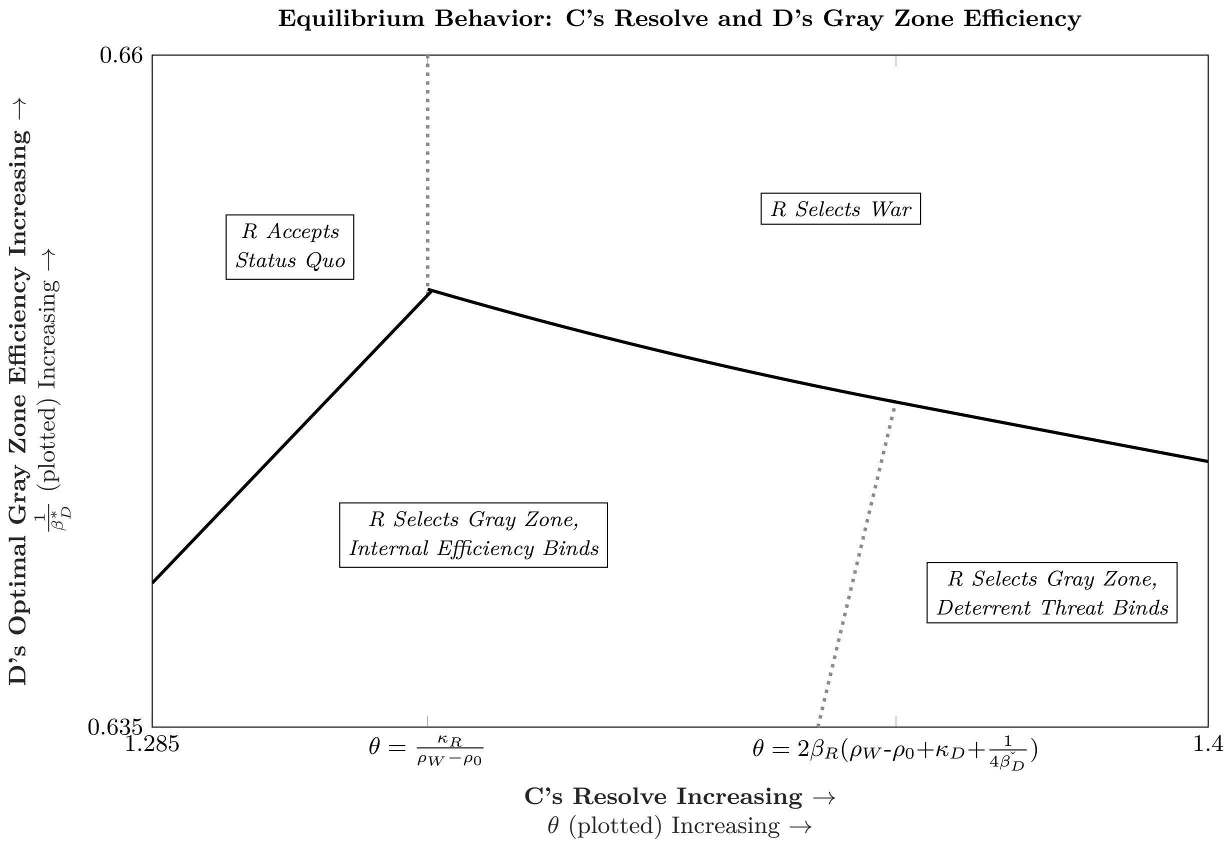


Figure A1: Extension 1: D's Optimal d^*

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 $\kappa_D = 0.1$.

1.5 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 $g_C > 0$, 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, 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 $w_R^* = 0$ 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 $g_D^* = \frac{1}{2\beta_D}$.*

[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.*

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

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

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 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. 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 $g_C^* = \check{g_C} = \frac{\theta}{2\beta_C}$. We can then express C's behavior in terms of θ. C prefers the status quo to gray zone conflict when

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

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

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

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.

2 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.

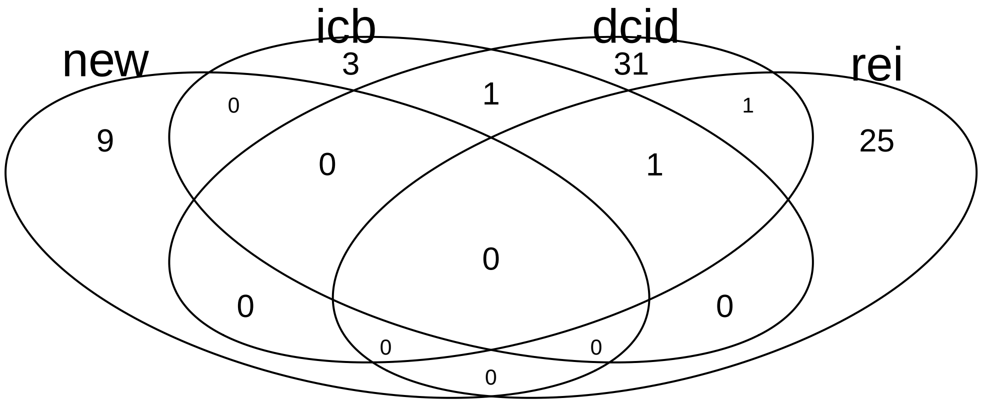


Figure A2: Venn diagram of case overlap among prior datasets

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 significant inconsistency concerning the sample of post-1994 Russian interventions identified by the ICB, DCID, and REI 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 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

Intensity of Russian cyber attacks (2005-2017)
Valeriano and Maness data

Intensity of Russian cyber attacks (1994-2017)
Way and Casey data

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

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 A1 present descriptive statistics of the sample used in the models provided in the main text.

Table A1: 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.

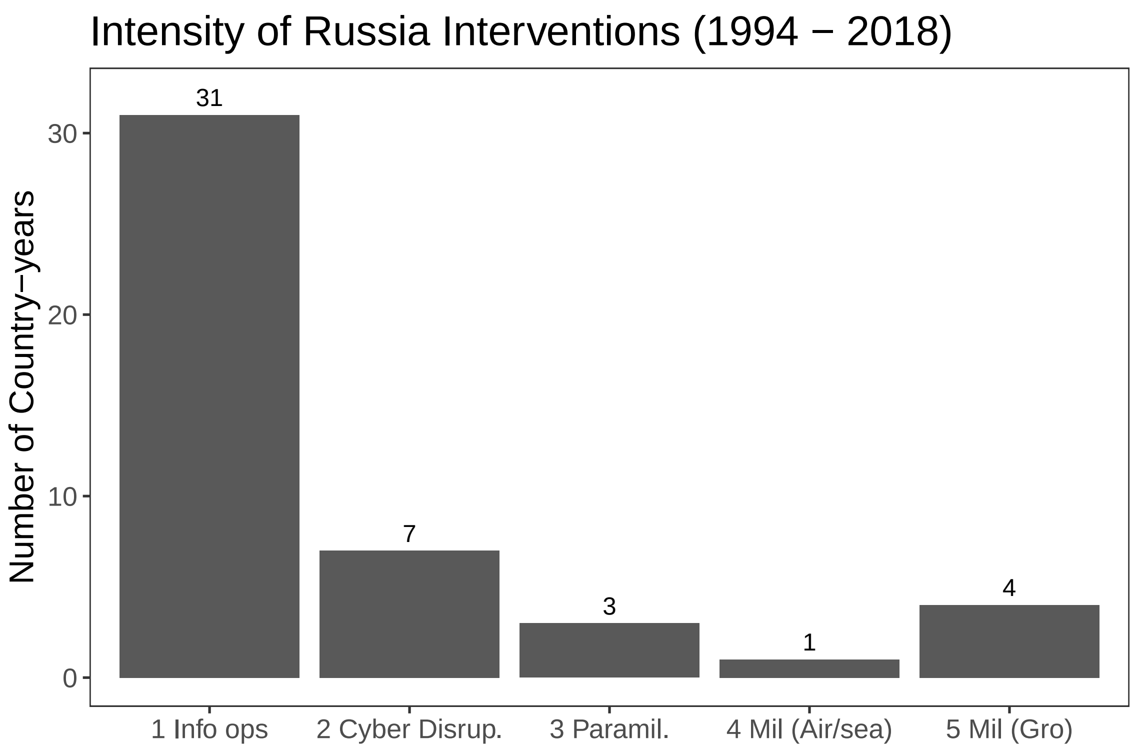


Figure A4: Intensity of Russian interventions

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. The intensity scale representing the dependent variable (the x-axis) appears to go to 6 instead of 5 because the ordinal value of 0 is numerically coded as 1, so all dependent variable values are shifted by one in this simplified depiction.

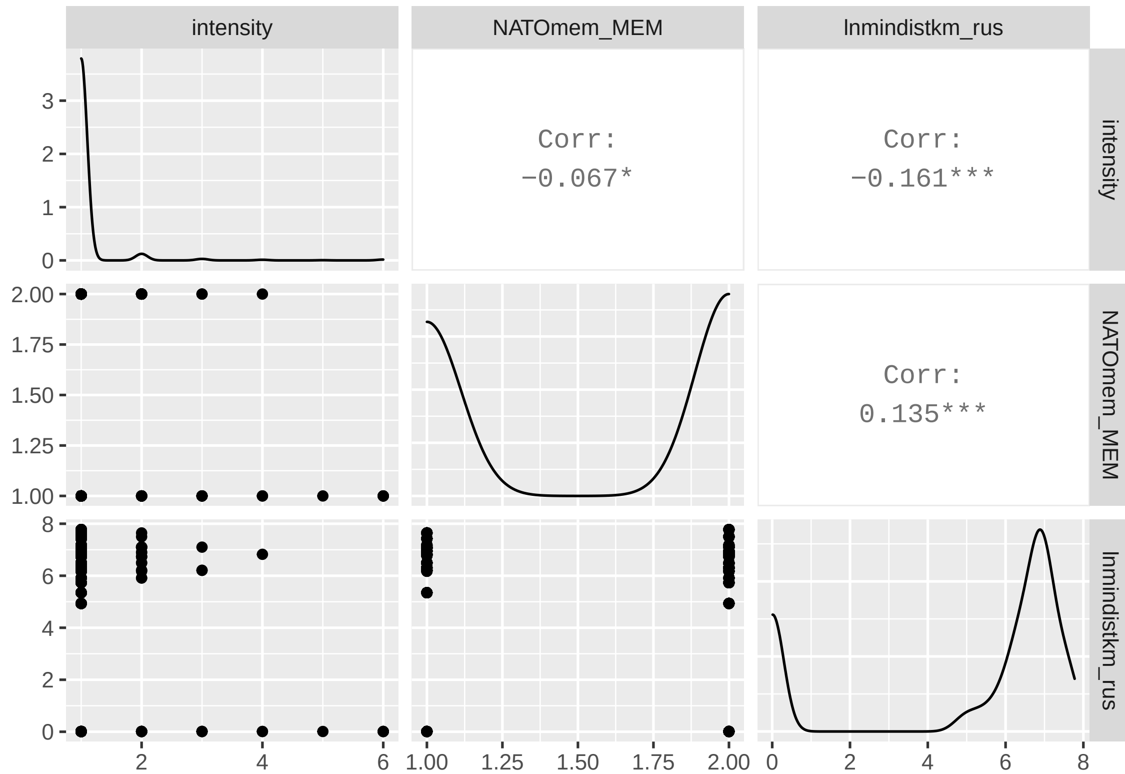


Figure A5: Bivariate correlation of dependent and independent variables

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 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.

3.2 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 A3.

	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*	0.57*	0.51*	0.42*
	[0.39; 1.12]	[0.29; 0.96]	[0.19; 0.91]	[0.15; 1.00]	[0.07; 0.94]	[0.11; 0.73]
Russia distance	0.90*	0.90*	0.88*	0.97	0.94*	0.97
	[0.84; 0.96]	[0.85; 0.95]	[0.83; 0.94]	[0.91; 1.03]	[0.90; 0.99]	[0.92; 1.01]
Controls						
Democracy		1.17	1.58		1.18	1.73
		[0.46; 1.88]	[0.89; 2.26]		[0.30; 2.07]	[0.83; 2.63]
Nuclear power		2.53*	1.56		1.59	3.79*
		[1.85; 3.21]	[0.84; 2.28]		[0.92; 2.26]	[2.61; 4.97]
Population		1.21*	1.15		1.15*	1.26*
		[1.06; 1.36]	[0.95; 1.35]		[1.03; 1.28]	[1.14; 1.38]
GDP per cap		0.99*	0.98*		1.00	1.00
		[0.98; 1.00]	[0.97; 1.00]		[0.99; 1.00]	[0.99; 1.01]
Mil. spending			1.02			0.97*
			[1.00; 1.03]			[0.95; 1.00]
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 A2: Intensity of Russian Intervention: Odds Ratios

	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^{***}	-0.22^{***}	-0.24^{***}	-0.32^{***}
	(0.02)	(0.03)	(0.03)	(0.07)	(0.08)	(0.10)
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		−0.05	0.06
		(0.12)	(0.13)		(0.20)	(0.22)
Nuclear power		0.12^{***}	0.09^{*}		0.09	0.12
		(0.04)	(0.05)		(0.08)	(0.12)
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	0.05
		(0.01)	(0.02)		(0.04)	(0.06)
Mil. spending			0.00			−0.00
			(0.00)			(0.01)
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 A3: Intensity of Russian Intervention: OLS Results

3.3 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 A4 are almost identical to those of the ordered probit, as expected.

	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^{**}	-1.23^{**}	-0.84	-1.28^{**}	-1.62^{***}
	(0.50)	(0.46)	(0.49)	(0.52)	(0.59)	(0.46)
Russia distance	-0.18^{**}	-0.20^{***}	-0.21^{***}	-0.04	-0.11^{*}	-0.02
	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)
Controls						
Democracy		0.79	1.27^{*}		0.67	1.30
		(0.82)	(0.76)		(0.98)	(0.91)
Nuclear power		1.70^{**}	0.97		0.68	2.81^{**}
		(0.82)	(0.88)		(0.79)	(1.36)
Population		0.48^{**}	0.42		0.37^{*}	0.60^{***}
		(0.22)	(0.28)		(0.19)	(0.21)
GDP per cap		-0.02^{*}	-0.03^{*}		-0.00	0.00
		(0.01)	(0.02)		(0.01)	(0.01)
Mil. spending			0.02			-0.07^{**}
			(0.03)			(0.03)
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: Intensity of Russian Intervention: Ordered Logit Results

3.4 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.

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 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 active civil war, 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 A5. We do not show results for models 1 and 4 separately to make clear that they models have no imputed values.

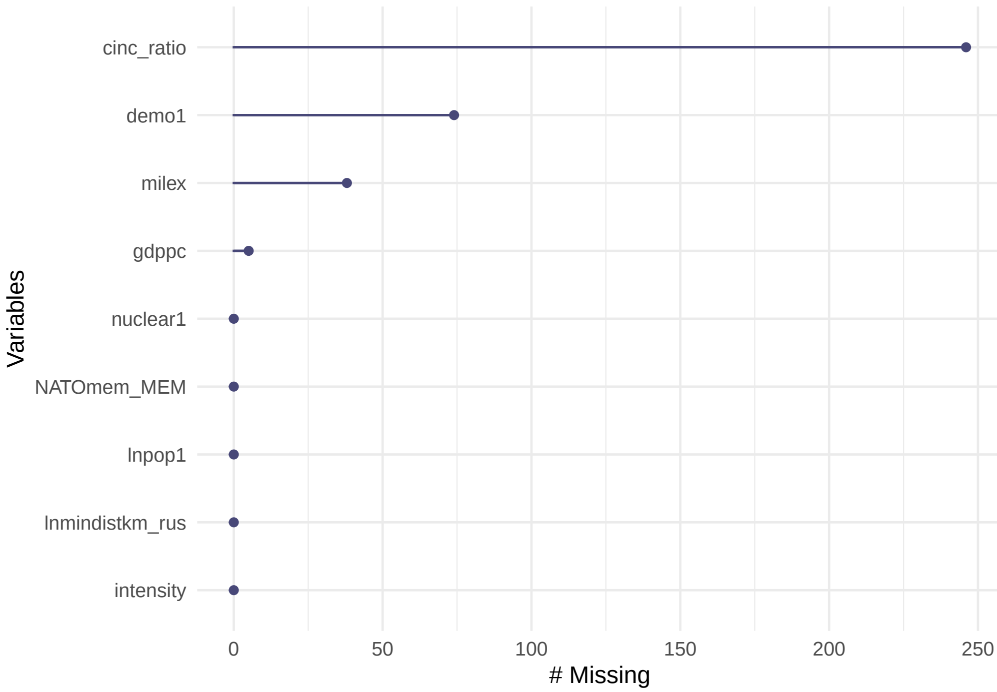


Figure A6: Number of missing observations for each variable in the dataset

The models in A5 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	0.57*
	[0.39; 1.12]	[0.15; 1.00]
Russia distance	0.90*	0.97
	[0.84; 0.96]	[0.91; 1.03]
Observations	1,000	376

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

Table A5: Intensity of Russian Intervention: Odds Ratios (non-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

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 ** (0.16 – 0.94)	0.37 ** (0.16 – 0.87)	0.28 ** (0.09 – 0.85)	0.29 ** (0.11 – 0.79)
Russia distance	0.81 *** (0.72 – 0.92)	0.81 *** (0.71 – 0.91)	0.90 * (0.80 – 1.00)	0.90 * (0.81 – 1.00)
Democracy	2.12 (0.44 – 10.36)	1.93 (0.45 – 8.29)	1.98 (0.32 – 12.32)	1.66 (0.29 – 9.60)
Nuclear power	5.61 ** (1.15 – 27.34)	3.70 * (0.79 – 17.29)	1.95 (0.43 – 8.80)	1.45 (0.24 – 8.72)
GDP per cap	0.98 (0.96 – 1.00)	0.98 * (0.95 – 1.00)	1.00 (0.98 – 1.01)	1.00 (0.98 – 1.01)
Population	1.58 ** (1.04 – 2.42)		1.46 ** (1.02 – 2.10)	
CINC ratio		732.46 *** (12.60 – 42575.63)		120.07 * (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)

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 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

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.

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.

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.

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.