Ref.: Ms. No. IO-6443

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

International Organization

Dear Mr Gannon,

We regret that we cannot accept your manuscript "The Shadow of Deterrence: Why capable actors engage in contests short of war" for publication in International Organization.

Comments from the Reviewers can be found below. As you will see, both reviewers express some interest in the model, but they are unpersuaded by the empirical evidence. While R1 thinks this is a potentially useful contribution to a current topic, the reviewer characterizes both the argument and the evidence as “obvious.” R2 recommends rejection on the grounds that the empirical analysis falls short of IO’s standards. Unfortunately, these reviews do not give us the level of support needed to move forward with this paper.

With respect to the theory, we agree with both reviewers that the model represents a solid effort to embed some intuitions about grey zone conflict into a formal logic. But the idea that deterrence at the strategic level can lead to conflict at lower levels, or policy substitution into lower cost (e.g., sanctions) or less easily detected (e.g., subversion) activities, is well known. Hence, the distinctive contribution is unclear, as is whom you are arguing against. The target seems to be policymakers who feel helpless to deter the low-level activities, and you are saying, “look on the bright side: the reason we are seeing low-level stuff is because we are doing a good job of deterring worse.” But if the concern is that we lack credible threats short of war for deterring grey zone attacks, nothing in the paper contradicts that.

As both reviewers note, the most interesting theoretical result is the material in section 7 showing that, if the defender gets better at countering grey zone attacks, this doesn’t necessarily make the defender better off, since that could push the adversary to war. But, as R1 notes, this depends on the conditions permitting a realistic chance of war. If the West got better at countering Russia meddling in its elections, would Russia really would resort to war instead?

With respect to the empirics, we share the reviewers’ concerns that the patterns uncovered are subject to multiple interpretations since a number of the model’s parameters vary with distance and NATO membership. Moreover, that Russia has reserved military action for non-NATO states near its borders, while doing cyber further afield, would likely be apparent to most observers.

As you may know, IO publishes only a very small fraction of the manuscripts that it receives. This means that we must sometimes decline to publish manuscripts that will eventually land in good journals. We hope the reviewer comments are useful in that regard.

We hope you find the reviewers' comments useful in your future research on this topic. Thank you for your interest in International Organization.

Sincerely,

Ken Schultz and Erik Voeten

for the Editorial Team at

International Organization

Comments from the Reviewers:

Reviewer #1: This paper argues that states engage in hybrid conflict, uses of force below the level of war, to shift outcomes in their favor while avoiding the risk of escalation to full-scale war. It presents a formal model of the choice to engage in hybrid conflict in the shadow of possible escalation to full-scale war. It then tests the conclusions of the argument against a data set of cases of hybrid warfare by Russia assembled from several existing data sets. It also provides short studies of illustrative cases. The argument is plausible; it strikes me as the obvious argument to make. The model does provide some interesting insights into the issue.

I have no problems with the model. It is straightforward given how it sets up the problem. It does what a formal model should do, clarify the problem through simplification so the logic of the argument is readily apparent.

I have more concerns with the data. The issue here is the empirical definition of what cases constitute "grey zone" conflict. The data is assembled from collections on cyber attacks, low level uses of force, international conflict, and election interventions by Russia in the last 25 years. While I understand the desire to be comprehensive in coverage of cases, the data lumps together events that may not match what grey zone conflict means in the model. In the model, it is an effort short of war to shift an outcome in one's favor when war is conceivable. But a number of the operations in the data occur when war is not on the table in any form. For one, the Estonia case of the cyber attack discussed as an illustrative case strikes me as Russia signaling its displeasure with Estonia rather than trying to change the outcome of joining NATO or moving the war memorial in question. The election interference cases also strike me as an issue where war is nowhere near the table

in any sense. They are not changing any issue that is the subject of international negotiations with Russia. Rather, they are efforts by Russia to interfere in democratic politics to discredit them generally and induce leaders less hostile to Russia. The results then devolve to the observation that Russia is more careful in its anti-Western acts further from its border and against NATO members. Yes, it is consistent with the model and its argument about hybrid warfare, but it is obvious even from casual observation. Part of the problem here is that the phenomenon is not well-defined but used in public policy circles without nailing down what it is.

Reviewer #2: This article has some real positives. It advances a basic idea -- that gray zone conflict reflects deterrence success -- which is an important rejoinder to the typical breathless view in DC that gray zone behavior by Russia/China represents a ghastly deterrence failure. The formal model is very clear and quite helpful in understanding the range of conditions in which gray zone emerges and when it reflects deterrence success. The issue of gray zone conflict is clearly top-of-mind to practitioners and touches on a range of topics (coercion; territorial disputes; cyber ops and covert action) that are both old and new areas of research.

The primary weakness, and the main reason I recommend rejection, is that the empirical component of the paper is not up to IO standards and unlikely to get there. I have a few several problems with the data and testing. Together they add up to a rather unimpressive empirical section.

--Start with the basics. The empirics address a single "user" of gray zone methods, in a limited number of cases, and for which there is significant problems of interdependence among those cases/observations (i.e. intervention tactics in 2008 reflect experiences in prior years). There is no sophisticated causal identification strategy. And the data, discussed next, are shaky.

--Sample under study. The authors (I assume plural from now on due to use of "we") admit that focusing on Russia's modern conflict activity requires relying on data that are "ambiguous." To be clear, this is because they have chosen to analyze a set of cases from the past 25 years (1994 to 2018) by a state that is non-democratic and notoriously guilty of misinformation. This means very little qualitative material will be available that address the main claims of the theory (perceived credibility of West's deterrent; reason for selecting gray zone and intensity thereof). It also means significant compromises in how variables are operationalized must be made (more below) which add to issues. As a basic matter, I am not sure how an article that focuses on this set of cases with these questions will generate credible empirics at the level IO expects. A different design that focuses on the past 25 years (such as qualitative with a government for which interviews and internal

material are available) could work. A different time period with Russia could work. But today + Russia is structurally stacking the deck.

--This is reflected in the one paragraph case studies that are stapled onto the end of the article, which are based on news accounts and add very little value besides simple illustration. The gold standard for qualitative empirics in a journal like IO is increasingly qualitative work that relies on excellent interviews or careful interpretation of primary evidence. These are far from this level.

--The article's central explanatory variable is strength of external deterrence ("we propose that the external deterrent threat from war is a key driver of Russian gray zone behavior"). The measurement of this variable is disturbingly and disappointingly blunt: a binary variable for membership in NATO. Readers will understand the intuition here. However, it seems a very coarse measure indeed, in two ways. First, operationalizing external deterrence strength this way does not address the range of factors beyond alliance membership that decades of work on deterrence and crisis diplomacy has identified as relevant to deterrence credibility. This includes military doctrine, reputation, private and public threats, and fine-grained military capabilities issues like location of deployment. None of these issues can be accommodated with a single blunt measure. Second, even assuming alliances are all we need to understand, a focus on "membership yes/no" omits many nuances about

alliances and deterrence identified in existing work by Leeds, Snyder, and more recently Poast and others.

--Separate issue: measuring "external deterrent threat strength" by NATO membership yes/no also creates an interpretation issue. Membership in an alliance does more than just boost deterrent. It also has knock-on effects on economic development, firm investment decisions, elite norms (think constructivist socialization work), and the consolidation of democracy in recently transitioned members. The results beg questions based on this ambiguity. Is Russia shying away from more provocative gray zone activity because of alliance-deterrence effects, or because of other political, diplomatic, normative effects of alliances? A natural way to explore this mechanism question would be to assess a set of cases over time where deterrence varies over time but, because alliance membership changes so rarely, that is impractical. Using a more refined measure of deterrence based on capabilities, or similar, would allow for more nuance and better exploration of mechanisms.

Also related to empirics but somewhat distinct are alternative explanations and selection problems with respect to these findings.

--As the authors note, it is tough to know what to make of the geographic distance / "gradient" issue. A lot more than deterrence changes as distance from Russia increases. Most importantly, the logistics of engaging in more kinetic forms of gray zone operations becomes very complicated the further from Russia one gets. Thus, resorting to cyber for distant targets may simply reflect practical considerations about which instruments of military power are viable, regardless of credibility of deterrence / alliance. Also very important: Russia's interests or sphere of influence are far lower the further from Russia's border. The authors note this (22) but these two issues suggest there are two alternative explanations that need to be carefully ruled out before we should give credit to external deterrence. Through quantitative or qualitative evidence, this manuscript has to convince readers that more mild forms of gray zone activity further from the "user" is \*not\* a result of

logistics or mild geopolitical interests.

--I worry about an earlier selection stage that could create a spurious correlation between NATO membership and the kind of gray zone tactics used. NATO membership decisions in the first place - who to admit and when - are deeply bound up in the credibility of Russian threats and perceptions of Russia's interests. Russia's own military capabilities and political rhetoric have likely deterred membership for countries like Ukraine that are also near Russia's border. What that means is that Russia's interests and capabilities play an important causal role in what the strength of the external deterrent Moscow faces when choosing gray zone tactics. This makes NATO membership appear causally significant. Russia-related variables may explain both NATO membership and gray zone tactic preferences.

My main recommendation is to change the article to focus on the formal model. Within that structure, I'd spend \*equal\* time on the two insights/observations from the model: a) strong external deterrence may explain gray zone and b) overly strong external deterrence can lead to rejecting gray zone in favor of outright war (on the manuscript's lack of attention to 'b', see below). I'd personally recommend shifting to illustrative case studies with a bigger dose of history (on this, see below) over the Russia empirics.

Smaller issues that may be helpful to the authors and editors.

--The authors' claim that gray zone warfare is not new could be said clearer. Little green or blue men used to revise the territorial status quo has been possible for decades. The Italians and Germans did something very similar in the Spanish Civil War in the 1930s. The addition of cyber and cyber-enabled information campaigns adds to the toolkit but arguably does not qualitatively change the availability of militarized actions short of war. I wonder if the paper could look at historical cases of "gray zone" activity (though it wasn't deemed such at the time) to obtain more empirical leverage than the focus on contemporary Russia allows.

--This historical connection does raise a puzzle of why there was no gray zone activity, at least under that heading, until relatively recently. A new spike in this kind of activity is, after all, why many in the DC/Pentagon see it as a response to some reduced deterrent.

--A final point based on the historical comparison. Should "gray zone" really even be the focus of this article? Their definition (the first sentence of conclusion section) makes it seem like gray zone is a synonym for "limited war." Scholars would certainly recognize the latter as long-running phenomenon in international security. So why use that label? Why not frame this as a model of limited war, then draw on the historical data available for those?