Ignore the Paradox: Russia, Ukraine, and Nuclear Deterrence
Written by Francesco Bailo and Benjamin E. Goldsmith

Ignore the Paradox: Russia, Ukraine, and Nuclear Deterrence

U.S. President Joe Biden's May 2022 essay in *The New York Times* refocuses attention on the debate over the risk of escalation to wider conflict, including nuclear war, in the current conflict in Ukraine. What useful tools do social scientists have to help understand these risks? Some experts writing in major policy magazines, newspapers, and online fora have made reference to the "stability-instability paradox." Is this widely known element of nuclear deterrence theory relevant to the current war in Ukraine? Our answer is "no, it is potentially misleading." The concept, introduced by Glenn Snyder in 1965, has for some become a standard assumption of nuclear deterrence theory. It proposes that rivals who each possess a nuclear deterrent may experience greater low-level conflict because the stability of nuclear deterrence will embolden one side to engage in lower-level conventional military action, confident the other side will refrain from responding decisively in order to avoid escalation to nuclear war.

In this article, we argue that the theory should not be given much weight by U.S. and NATO policy makers, because it lacks factual support. Rather, we suggest how discarding the paradox can lead to different insights and policy guidance based on well-supported concepts from conflict and deterrence studies.

At first glance, the stability-instability paradox seems a good description of the current Russia-Ukraine war. Indeed, Andrew Kydd makes the case that "the stability-instability paradox actually gives hope that escalation may be avoided, rather than additional reasons for fear." If Russia, the weaker conventional power, has been emboldened to invade Ukraine, this is likely due to confidence in a restrained U.S. response, consistent with the paradox's prediction that nuclear states will not escalate beyond "low-level" conventional conflict, such as a proxy-war.

However, Emma Ashford and Joshua Shifrinson also employ the paradox to make a different argument. They claim that, under "the stability-instability paradox, in which states, stalemated in the nuclear realm, might be more willing to escalate in conventional terms," there is a greater chance of "escalation to a broader [conventional] war." Among "multiple paths" to potential wider conflict they point to the increased presence and proximity of NATO and Russian troops (including in Belarus) which make "accidental confrontation between the sides more likely," and the possibility of a Russian attack on Western weapons supply lines. Kydd disagrees, arguing that "while the paradox frees up Russia to invade Ukraine, it also frees the West to respond with aid to Ukraine, so long as that does not cross the threshold to direct engagement with Russian forces" (moving the conflict from low- to medium-level).

Before rushing to apply the paradox for policy prescriptions in this exceptionally dangerous environment, let's consider three fundamental questions. First, how much stock should be placed in the stability-instability paradox? Second, do we need the paradox to explain the behavior of Russia and the United States in this crisis? Third, what might be the benefits of ignoring the paradox?

Where's the evidence?

To address the first, we acknowledge the paradox's long and distinguished theoretical pedigree, including Snyder and other major scholars such as Charles Glaser and Robert Jervis. Both Frank Zagare and Robert Powell developed game-theoretic models demonstrating its logic. However, a crucial caveat has so far been ignored: factual

Ignore the Paradox: Russia, Ukraine, and Nuclear Deterrence
Written by Francesco Bailo and Benjamin E. Goldsmith

evidence for the existence of the paradox is at best equivocal, and indeed in our view very flimsy.

While the idea is intuitive, it has not fared well in recent empirical studies. The latest large-sample statistical study (2015), by Mark Bell and Nicholas Miller, finds pairs of nuclear states are not "significantly more or less belligerent at low levels of conflict." Using data for the period 1945-2000, their results,

[do] not support the stability-instability paradox …. Nuclear dyads do not appear to behave differently to nonnuclear dyads—neither more or less likely to fight wars nor significantly more or less likely to engage in low-level conflict than nonnuclear dyads.

Our own recent (2021) empirical research also provides at best mixed and weak evidence for the paradox, using a new method for quantitative comparative case studies, and 1954-2010 data. We conclude: "Our results cast doubt on the theory, even after we give greater precision for its applicability. There appears to be little systematic empirical support that is not vulnerable to claims of bias due to selection on the dependent variable, extrapolation from irrelevant cases, or fraught interpretation of indeterminate qualitative evidence." Crucially, the only case for which the evidence fits the theory is India-Pakistan during the Kargil War. We point out that the experts on this conflict are far from a consensus that the paradox actually applies, while the evidence employed in the debate is indeterminate. If the endurance of the theory itself is largely due to this case, the inferential fallacy of selection on the dependent variable is likely to blame, distracting from valid alternative explanations.

In short, the theory lacks compelling support in the historical record of militarized confrontation between nuclear states. Not every elegant or logical social science theory will be "true."

The right theory?

Moving to our second question, do we need this theory to interpret the current crisis? Our answer is that it is not very relevant or informative. The question the paradox poses is not whether nuclear deterrence will hold, but whether Russia is exhibiting emboldened behavior due to its *belief* that nuclear deterrence will hold. It is primarily a theory about conventional conflict between nuclear powers. If we want to ask: would Russia have invaded Ukraine if it and the United States did not possess nuclear weapons, then the stability-instability paradox is a relevant theory. And it would suggest the answer is "no."

We also believe the theory is too vague to give specific, meaningful policy guidance. Consider Kydd's claim cited above that the paradox frees up the West to provide military support to Ukraine "so long as that does not cross the threshold to direct engagement." We are not aware of any explication of the theory providing such clear guidance for where to draw the lines between low, medium, and high levels of conflict. There is no such indication in Snyder's original statement of the theory, nor in Powell's recent game-theoretic elaboration. Kydd's assessment strikes us as *ad hoc*. And indeed both he and Powell reference the 1999 Kargil War between India and Pakistan as a (perhaps *the*) key example. The problem for Kydd's argument is that this was undoubtedly a case of "direct engagement" between two nuclear powers. Similarly, Robert Litwak implies that the theory is relevant when peripheral interests are involved, but not interests "vital" to the nuclear states. We are aware of no such distinction in the relevant literature.

If the focus is on the risk of escalation to nuclear war, we believe the relevant theory is simply that of nuclear deterrence. While Bell and Miller, as well as Erik Gartzke and Matthew Kroenig, point out that less than commonly assumed may be well-understood about the impact of nuclear weapons on international security, the evidence to date regarding nuclear deterrence and mutual assured destruction (MAD) is clear – there are no instances of nuclear war between two nuclear powers.

So when considering the risks of NATO involvement in the Russia-Ukraine war, current social science does not point to the stability-instability paradox for guidance. Rather, it points to two more established areas of knowledge. First, that of nuclear deterrence. The best evidence we have is that nuclear deterrence will hold, as it did during the Vietnam War, other Cold War crises, and India and Pakistan's Kargil War. The existential logic of MAD should be hard for policymakers to ignore. Second, it points to the extensive knowledge we have on conventional conflict and

Ignore the Paradox: Russia, Ukraine, and Nuclear Deterrence
Written by Francesco Bailo and Benjamin E. Goldsmith

deterrence. That conventional escalation may lead to nuclear escalation is a long-standing assumption, but there is little evidence that this escalatory ladder is well described by the paradox.

Just like any social science theory, the stability-instability paradox should clear some serious hurdles of empirical testing before we place faith in it for policy prescriptions, or accept it as established knowledge. So far, while it may have a catchy name and an intuitive argument, it has not passed this test.

Three Reasons to Ignore the Paradox

Even if it lacks empirical support, isn't it wise to consider the paradox given its intuitive logic? We argue that there are at least three ways in which giving undue credence to the theory might distort thinking about the current crisis.

First, the unsubstantiated and sometimes contradictory guidance based in the stability-instability paradox can be discarded. For example the confusing advice of Kydd that any direct conflict is to be avoided because the paradox considers this a clear escalation – when in fact there was direct conflict between India and Pakistan during the Kargil War – can be set aside. We should evaluate the risks of escalation due to direct conventional fighting between nuclear states based on the logic of MAD, without the complicating and unsupported assumptions of the paradox. For example, despite some robust rhetoric, Russia has fairly clearly and consistently signalled that only an existential threat to the Russian state would lead it to consider using nuclear weapons, and neither the war in Ukraine nor the consequent possible NATO expansion to Sweden and Finland poses such a threat.

Second, it frees us to return to the essential elements of nuclear deterrence theory considered specifically in the context of this very novel confrontation between a diminished Russia and a United States simultaneously considering the rise of another nuclear superpower, China. In an earlier study, Robert Powell reminded the field of the continuing importance of Thomas Schelling's insight that coercion under MAD relies on "threats that leave something to chance," even in the post-Cold War era. The exact nature of this brinkmanship contest should be considered in the current context, with the fundamental trade-off for the U.S. being how much coercion to attempt (e.g., how robustly should Ukraine be supported) while avoiding being out-bid in brinkmanship signalling by Russia. The paradox also tells us nothing about China's role, either in terms of current support for Russia, or in terms of how U.S. handling of the current crisis might affect its capacity to deter China from attacking Taiwan in the future. But these are crucial issues that deserve focused attention, potentially as a 3-actor brinkmanship model. One way that the U.S. and NATO might signal this element of chance to China (and Russia) is to acknowledge that it is contemplating direct military involvement, for example either in response to revelations of atrocities or in order to ensure Ukrainian "victory" that Western leaders now openly advocate. Would the U.S. actually join a conventional fight to repel a Chinese invasion of Taiwan? The more Chinese leaders contemplate this as a genuine possibility, arguably the stronger U.S. extended deterrence will be in the future (although China might also strengthen its own capabilities in response). Systematic strategic thinking is developing to address this sort of emerging nuclear complexity.

Third, discarding the paradox as a reliable guide allows us to focus on what we do know about rational conventional deterrence and conflict. How can Russia best be deterred from widening the conventional conflict, either (again) beyond the Donbas in Ukraine, or to new fronts such as Moldova or the Baltics? While the assumption of the stability-instability paradox is that the nuclear balance can embolden a weaker nuclear state, the assumption of conventional deterrence *leads to the opposite prediction* – the weaker state will be cautious because of the conventional balance. Since there is no systematic evidence of such emboldened behavior, is it wise to build this assumption into our understanding of Russia's current and future behaviour in this crisis? We think not.

Indeed, Russia seems to be treading very carefully, even while its conventional war seems to be failing to meet its objectives. Increasing the conventional military pressure – reducing expected benefits and increasing expected costs for Russia – while also taking care that Putin has a viable off-ramp in his domestic political context (as much as this is possible) might deter expansion of the war and also compel a faster end, given that Russia may be planning for a long war of attrition. Such an approach seems reasonable given the current evidence, and is consistent with the provision of advanced Western weapons systems, but it is not consistent with the stability-instability paradox's assumption of an emboldened Russia leading to the necessity of restraint from the U.S. and NATO.

Ignore the Paradox: Russia, Ukraine, and Nuclear Deterrence

Written by Francesco Bailo and Benjamin E. Goldsmith

In this short analysis, while we provide what we believe are logically and empirically well-supported policy options resulting from ignoring the paradox, we certainly do not claim to examine all implications of a robust conventional deterrent strategy towards Russia. But we do claim that there is no systematic empirical evidence that the assumption of emboldened and risk-acceptant conventional conflict behavior of Russia is in some way caused by, or protected by, MAD. Whether Russia will back down or escalate to robust conventional signalling is a question that should be considered on the merits of the current situation and well-supported theory, not the claims of the stability-instability paradox.

About the author:

Francesco Bailo is a Lecturer in the School of Communication, University of Technology Sydney. He is a political scientist / data scientist with broad experience, including working for the United Nations, and is author of *Online Communities and Crowds in the Rise of the Five Star Movement*.

Benjamin E. Goldsmith is a Professor in the School of Politics & International Relations, Australian National University. He has been writing in leading journals on interstate conflict and security for two decades. Recent articles have appeared in the *American Political Science Review* and *World Politics*.