Strategic Stability in a Rapidly Changing World

Sheena Chestnut Greitens: Welcome to *Horns of A Dilemma*, the podcast of the *Texas National Security Review*. I'm Sheena Chestnut Greitens, editor-in-chief of *TNSR*, and I'm here today with Dr. Ryan Vest, our executive editor. We're pleased to have Harold Trinkunas join us today, who, together with Herb Lin, recently put together a special edition of the *Texas National Security Review* focused on emerging technologies in the future of strategic stability. Harold is the Deputy Director of the Center for International Security and Cooperation, and a senior research scholar at the Freeman Spogli Institute for International Studies at Stanford University. Harold, welcome to *Horns of a Dilemma*. It's great to have you on the show.

Harold Trinkunas: Thank you so much for having me on the podcast.

Ryan Vest: Harold, your article is an introduction to a special issue that looks at emerging technologies and strategic stability. In it, you argue that many of our assumptions about deterrence were shaped by a specific historical era, the US–Soviet nuclear standoff. What assumptions from that era no longer hold in today's technological environment?

Harold Trinkunas: Thanks for the question, Ryan. I'd say that the period of US–Soviet strategic stability that lasted from about the 1950s through the end of the Cold War was really shaped by a very specific relationship between the two powers. It was a bilateral relationship. Both sides had large numbers of nuclear forces.

They both had secure second-strike nuclear forces, and that created a number of incentives for the both powers to avoid escalation during the period. What's changed today, the scholars in this Roundtable argue, is really three things. One is, there's a larger number of players that have nuclear weapons, making the relationships among them less predictable.

Specifically, at least US, China, and Russia are in a nuclear standoff, if we want to call it that, that is proving particularly intractable to solutions such as arms control, which once made the US Soviet relationship more predictable and frankly safer for everybody on the globe.

Another force that's changed is that the emerging technologies we discuss in the Roundtable, and we focus on things such as the employment of artificial intelligence, of cyber offense and defense weapons, hypersonics, is that these technologies are available to a wider range of powers. There's been horizontal proliferation in a sense of these emerging technologies, and even middle powers, such as, for example, the Netherlands, can achieve very sophisticated cyber capabilities, for example. So this, of course, complicates the ability to maintain a stable and predictable relationship.

Another thing that we find in the Roundtable is that the combination of emerging technologies can allow states to achieve effects that once were only thought possible with using nuclear weapons.

For example, widespread sensors combined with artificial intelligence powered integration and analysis of data creates the possibility that at some point in the near future, you might be able to achieve a much more sophisticated intelligence picture of where adversary nuclear forces are located, which of course might allow you to put their second-strike capabilities, such as on submarines or mobile platforms, at risk.

You can imagine that cyber can be used to turn off air defenses and otherwise make states more vulnerable to a first strike. And the final thing that we look at, and I think is a particular feature of this Roundtable, is the role of social and individual psychology, particularly in leader psychology.

And we really see this in two different dimensions. One is a short-term dimension, which is time compression to make decisions about nuclear weapons employment during crises. But we also see the possibility of long-term effects that might appear early in the acquisition cycle for new weapons.

We see overconfidence and over-optimism being an issue as both leaders and weapons designers and acquirers engage with these emerging technologies, and, on the whole, that also makes the system less predictable.

Sheena Chestnut Greitens: So, as you mentioned a minute ago, the Roundtable really covers a wide range of emerging technologies from AI and machine learning to things like cyber, space, advanced conventional weaponry, and, then these psychological factors that I do want to come back to in a moment, but I wanted to ask you whether there was a particular technological development or moment over the past few years that clearly signaled to you that old assumptions about strategic stability might not hold, and that the whole conversation might be entering a new and more dangerous phase?

Harold Trinkunas: So I'd actually say that it emerged, in my case, and I also think this is true for my colleague, Herb Lin, that we together co-edited a volume several years ago called *Three Tweets to Midnight* (*Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*), which looked at the possible risks associated with the modern information ecosystem dominated as it is by social media and new media on nuclear crisis stability.

And we saw incidents from that which raised the question of how other emerging technologies might have an effect on crisis stability. We looked at everything from Pakistani defense officials tweeting nuclear threats to Israel in response to a fake tweet to the use of social media to raise questions for US forces in Korea as to whether their family dependents needed to start evacuating.

And that made us just wonder what other emerging technologies might have effects that were new and potentially destabilizing. So, that was then accentuated by the outbreak of the war between Russia and Ukraine, which has really foregrounded the use of a number of these systems, and on the one hand raised the question of instability— and we're are all aware of President Putin's nuclear saber rattling—but also of how non-nuclear powers like Ukraine might be able to use these emerging technologies to better defend themselves and bring their different calculus of risk and reward to these kinds of conflicts then we might've expected before the outbreak of that war.

Ryan Vest: That's really interesting. You know, one of the key arguments, one of the core arguments of this issue, as I was reading through it, is that emerging technologies are changing the calculus of nuclear doctrine and decision-making in ways that can undermine strategic stability. I was wondering if you could talk a little bit about some of these ideas and what you mean by undermining or what these new technologies are, and why you see this as an essential problem of the Roundtable to address.

Harold Trinkunas: Sure. I think there are really two different sets of scenarios. One is the calculated use of nuclear weapons. And the second is accidental use or inadvertent escalation. And I think the first thing we might see as emerging from simply the emerging technologies giving one nuclear-armed state an advantage in a crisis that it might otherwise not have had in the past. For example, maybe its ability to use cyber offensive weapons to interrupt their adversary's nuclear command and control systems.

But a second case is this issue of accidental use or inadvertent escalation, which lies at the center of a different set of problems associated with automation. And in many cases, artificial intelligence, for example, is introduced into systems—what we're seeing is very advanced forms of automation. Of course, we're all I think familiar, or some of us are familiar, with past instances where automation set off the possibility of accidental or inadvertent use of nuclear weapons.

These are, for example, the training tapes incidents at NORAD in the late 1970s, where training tapes simulating a nuclear attack were accidentally loaded, and human operators initially thought they were seeing a real attack on their systems. In a similar incident in the Soviet Union in the early 1980s, as well, it was only the intervention of the human watch officer that prevented the Soviets from escalating in response to what they perceived as an incoming US attack.

So, I think those are both scenarios in which emerging technologies could possibly either change the strategic calculus if they put second-strike forces at risk or create a risk of bad information coming into the system and affecting decisions for human operators.

But we also look into the psychological dimensions of this as well. And particularly Rose McDermott in her piece (*The Influence of Psychological Factors in the Search for Strategic Stability - Texas National Security Review*) does, I think, a good job of explaining why emerging technologies can affect human decision-making by undermining the rationality, or the ability to make rational decisions, of leaders.

One effect that she looks at is the effect of time compression. And we know from the work of Daniel Kahneman, who wrote the book *Thinking, Fast and Slow* (Daniel Kahneman-Thinking, Fast and Slow .pdf), that humans tend to use heuristic biases to make decisions when working under pressure. And these are sort of your gut responses, your emotional responses, which are not necessarily wrong, but they're certainly not rational.

And the second concern she raises, and she does talk about a number of other heuristic biases that, for example, overoptimism among leaders, overconfidence as well. But there's this particular issue that she raises today, which is that we are increasingly in a world led by personalist regimes, and that's true in Russia, it's true in China, and it's true in the United States. And she also points out that many of the countries seeking to proliferate in the nuclear direction are also led by personalist leaders.

So this is a way in which psychology has an effect on the probability of nuclear risk. And where emerging technologies intervene is by compressing the timelines on which leaders have to make decisions and therefore activating very likely that fast thinking, working from the gut, working from instincts to respond during crises.

Sheena Chestnut Greitens: I find this really fascinating, and I want to come back and sort of dig deeper into the chunk of articles in the Roundtable that deal with these psychological and organizational biases and factors that take us away from what we think of as very rational decision-making when it comes to nuclear deterrence. But before we do that, I wanted to ask you about something before I forget to do it. I want to ask you about something that you said in the Roundtable, and that is, throughout a couple of the pieces in it that really jumped out at me the first time I read them. And that is this central claim that you mentioned a minute ago, that emerging technologies can now create these effects that were previously reserved for nuclear weapons.

This is, in some ways, really counter to the way that I think many of us think about nuclear weapons, which is that they are qualitatively different and a category apart, but the Roundtable really challenges that and walks through some ways in which emerging technologies erode the sort of separate logic of nuclear deterrence.

And so I wonder if you could walk us through what the Roundtable tells us about the blurring of this distinction, about the way that emerging technologies achieve these effects, and fundamentally why that matters for strategic stability and for deterrence and escalation risks.

Harold Trinkunas: Absolutely. So I think that where this comes in is really in at least two dimensions. And this is again going back to the idea of strategic calculation leading to the use of nuclear weapons. One is the ability to pinpoint and successfully attack an adversary's second-strike nuclear force is seemingly becoming more of a real-world scenario as sensors are more widely deployed, as overhead imagery becomes more widely available, and as artificial intelligence-powered integration analysis is increasingly used in military systems.

So this might create the temptation that you could have so-called exquisite situational awareness and know exactly where all your adversary's second-strike forces are. And that could tempt leaders to try to take the chance that they can take out all of their adversary's nuclear forces. That's one change, but a second

change is a possibility that you could do it without having to use nuclear weapons.

In the past, it was always assumed that to take out, for example, an adversary's land-based missile silos would require the use of nuclear weapons to achieve the kind of blast that would successfully destroy those missiles before they were launched. Now, people think that using some sort of conventional, advanced conventional weapons and precision strike, you might be able to achieve that effect, which was once thought to be reserved solely for nuclear weapons. That doesn't take away the unique quality of nuclear weapons, for example, in terms of how it affects, you know, large population centers or the long-term lasting effects of their use subsequent to their employment.

But at least in terms of being able to get at your adversary's nuclear weapons and destroy them, it does create the possibility that you could do that without you yourself having to use nuclear weapons. And I think that was one of the scenarios that really drove some thinking in the Roundtable at the beginning as an example.

Ryan Vest: I'd like to come back around to something you mentioned just a minute ago when you were talking about some of Rose McDermott's work. Throughout this piece and throughout the Roundtable, you highlight three major ways that emerging technologies can undermine strategic stability. You talk about incentives for preemption, horizontal proliferation, and psychological effects on decision makers. I was wondering if you could go through and expand a little bit on each of these in more detail and explain which of these worries you the most and why.

Harold Trinkunas: Interestingly, I think that psychological dimension is the one that most concerns me, especially as we combine it with the change in the nature of the regimes that govern, you know, that preside over the great powers, the nuclear-armed powers. And the fact that you have highly personalist regimes, you have uncertainty around crises because of the simply unknown effects of the interactions of emerging technology systems. And the time compression uncertainty and overconfidence and overoptimism in leaders is, I think, a dangerous cocktail, where we can see people having to make decisions under short timelines in unpredictable circumstances, different than the bilateral relationship that might've once existed between the US and the Soviet Union.

Now, in terms of horizontal proliferation, I think this is where we're starting to see things such as the effects of drone technology, the possible development of hypersonics, and the widespread development of precision strike.

We're seeing just how fast technology can develop in these areas and, for example, the Russia-Ukraine War. The Ukrainians have just accelerated drone development to an unprecedented extent and also acquired long-range strike capabilities, cruise missile capabilities, sea drones. I mean, this conflict, it really, you know, drove for us the importance of understanding how emerging technologies might actually interact on the battlefield, because it's unexpected.

I think one of the things that, for example, in his piece (*Technological Surprise and Normalization Through Use: The Tactical and Discursive Effects of New Precision-Strike Weapons in the Russo-Ukrainian War - Texas National Security Review*) Cameron Tracy tries to do, is draw out what were the initial expectations of advanced conventional weapons systems prior to the conflict's outbreak, and then looks at how they actually performed in the conflict itself.

But I would also say that, strangely, the war reduced some of the pessimism that some of the Roundtable authors initially felt about emerging technologies. Because it showed that, and I think this is important, just as emerging technologies can allow more countries to achieve more effects that they weren't able to achieve before, they also allow for the emergence of counters to the use of emerging technology.

So just as, for example, we can imagine that the combination of AI and widespread deployment of sensors would lead to better intelligence on adversary's second-strike forces. Similarly, we can envision data poisoning of the training sets that military AIs use to develop the capabilities to identify adversary systems, or simply disinformation, or putting false artifacts out into the environment so it becomes harder and harder to track these forces.

Obviously, it's still an evolving scenario. We don't know exactly how this is going to end, but I think the fact that you can develop counters to the effects of some of these emerging technologies at least brings up the possibility that long-term we could return to some form of stability when it comes to strategic stability.

Sheena Chestnut Greitens: I wanted to ask you a question about the nature of grappling with emerging technologies and their effect on modern warfare. Three of the essays, at least three, maybe more, but three that are jumping to the top of my mind here, all deal with this potential for there to be a real mismatch between what people think weapons are going to do to shape warfare, and then the actual effect that they have.

So Herb Lin's piece (*Artificial Intelligence and Nuclear Weapons: A Commonsense Approach to Understanding Costs and Benefits - Texas National Security Review*) talks about techno-optimism in the modern American defense technology sphere. Cameron Tracy's piece, which you just mentioned, talks about the idea that whether we over or underestimate the effective technology, there's a normalization process that calibrates expectations and adjusts them over time. And then Mike Horowitz, in his piece (*Artificial Intelligence and the Future of Strategic Stability - Texas National Security Review*) unpacks this non-linear process of overestimating and then maybe underestimating the transformative effect of emerging technology.

So each of these pieces takes a different approach to assessing how militaries, combatants, political decision makers adjust to emerging technology and figure out how it's going to shape warfare. I wanted to ask you whether there was anything that surprised you about that process, that was unexpected that you uncovered in the process of writing the special issue, or what do you think maybe it's most important for decision makers to know about this uncertainty as we try to assess the effect of emerging technologies.

Harold Trinkunas: I think that's a fantastic question. And I think going into the project, when we first started organizing the Roundtable, our focus was much more on the social psychology and individual psychology and heuristic biases—so decision makers in the moment during a crisis, and what potential impact emerging technologies could have there.

But what Mike, Herb, and Cameron all highlight is the long-term impact of psychological biases, of individuals making decisions and defense establishments, institutional biases towards acquiring and deploying different kinds of weapon systems, and how to integrate them into military operations.

And that was something I didn't really expect going into the Roundtable, that the psychological effects were much more widespread within the system than just as focused initially on individual leaders. And in a sense, because we were concerned about strategic stability and the possibility that escalation might lead to either calculated or accidental use of nuclear weapons,

you, of course, focus on those leaders making the decision at the moment. The leaders, which in the case of the United States, just the president, in the case of other great powers, might be a very small group of individuals, are making that decision. And I think that gave us too much of a short-term focus as to the impact of psychological factors and not enough focus on broader organizations and their psychology.

And so I think something that all three authors did very nicely was broaden out and lengthen the time horizon that we were considering of how human psychology affects decisions about acquiring and deploying emerging technologies.

Sheena Chestnut Greitens: It reminds me a little bit of the Thucydides line "War is a stern teacher," right? But teaching takes time.

Harold Trinkunas: Yes.

Ryan Vest: So, Harold, you discuss risks like automation bias, overconfidence, and reliance on AI-enabled decision aids. Is the danger here more about bad data, misplaced trust, or the speed at which decisions now have to be made? What do you think?

Harold Trinkunas: Well, I think all three are factors. One that I think we don't think enough about is bad data, especially because the training of artificial intelligence models requires large amounts of data, and oftentimes we're talking about unknown battlefield situations, unknown conflicts, things that really haven't been witnessed before or have happened very few times in history, such as decisions about nuclear weapons use.

Others, such as automation bias and excessive trust in the system, are things that we're maybe more familiar with. We've already had to go through adaptation to these in previous generations of military systems. Especially when we think about things such as automation—automation has been a thing on the battlefield for a very long time now, at least since World War II, and so military organizations have had to adapt their training and organization to try to mitigate some of the risks associated with automation. So I place less of an emphasis on that. But the final fact you raise about speed of decision making, and oftentimes in the military, there's this emphasis on trying to get inside their adversary's decision loop, so to speak. Being able to make decisions more quickly than the adversary, of course, starts to then trigger the psychological effects of thinking fast, thinking using heuristics, and instinct–which, I think, has its own set of risks of miscalculation or accident, and as well as just poor calculation.

Sheena Chestnut Greitens: I was struck by the emphasis here on how emerging technologies have kind of two main effects. One is the one you just mentioned, which is this compression of the timelines allowed to decision makers for decision-making. And we often think of the ability to decide and execute quickly as an advantage, but a number of the pieces in this Roundtable really point out that slowing down and allowing for deliberation, that Type Two

thinking, I think it is, can really actually help decision makers make more rational or better decisions.

And the other is just a fundamental uncertainty that exists around the effects of these emerging technologies, which was the point I was asking about a minute ago—the fact that a lot of our initial assessments of how emerging technology might impact warfare seem to be inaccurate or at least end up getting adjusted as decision makers gain experience with these weapons. So, I wondered if you could talk a little bit about how you see that amplifying decision makers' cognitive biases, and what effect does that have on strategic stability?

Harold Trinkunas: So I think there's a couple of different ways to think about this. I think uncertainty around the effects and interaction between emerging technologies and existing technologies, such as nuclear weapons, means that it's much more difficult to successfully plan ahead, train, and predict how a crisis might unfold.

And so one way you might manage risk during a crisis, even if you have to make decisions quickly, is by setting up a process that includes deliberative steps that would have to be taken before a decision is made involving the use of nuclear weapons. And clearly, for example, we're familiar with that on the US side. And in fact, I think recent films have actually even looked into this as well. But the uncertainty and the unexpected that might happen in a crisis, because we're simply not familiar with how emerging technologies might be employed, again, I think, triggers thinking fast among leaders.

They have to use their gut; they have to go with their instinct. They've never seen the situation before. And when that leads to decisions associated with escalation, I could see it being particularly dangerous. Rose McDermott, in her piece, does talk about a number of heuristic biases.

I don't want to focus too much on time compression, but there's also this sense that leaders might not really comprehend the scale of what they're deciding. She talks about psychic numbing, which is an effect that Paul Slovic and some of his colleagues have posited with relation to decisions about national security that involve the possibility of large numbers of casualties, simply because both leaders and ordinary people just cannot conceive of casualties of the scale involving hundreds of thousands or millions of people. And so it becomes a very abstract decision at that point that could also factor into leaders making decisions that have enormous consequences without fully thinking them through.

That said, one of the places in the Roundtable we do look for some optimism is: are these emerging technologies such that you could lengthen timelines for decision, have better information earlier on, better analysis, help leaders have just more time to engage in Type Two thinking about crises?

And, that's where we hope to see more work in the future, is to think about is there a way to give people more time to decide?

We've had to deal with emerging technologies in the past, and Frank Gavin has a fantastic piece in the Roundtable (*Strategic Stability and Its Limits: Reflections on Schelling - Texas National Security Review*) where he talks about how many of the emerging technology questions we're struggling with today were first raised in the context of integrating nuclear weapons into military establishments. Questions of uncertainty, questions of surprise, questions of time compression, and inability to really successfully conceive of the scale of destruction that would be possible. So I think just the key ingredient of time, time and experience with these emerging technologies, is an important factor.

So, for things to go right, you need to have enough time to come up with approaches that restore stability to the international strategic environment. And of course, we are seeing some efforts in that direction. For example, the decision by Presidents Biden and Xi Jinping to say that artificial intelligence would not be used to make decisions about nuclear weapons employment is an example that the leading powers are starting to focus on this problem. Of course, there is an issue with that statement in that AI is already used everywhere to the left of the decision on nuclear weapons employment, even if maybe it's not used in the final decision.

So I think we're still going to struggle with this, but time has the effect of solving many of these problems as humans gain experience with the possible interaction effects among the emerging technologies. And in fact, you may see a bit more caution going forward as leaders realize increasingly that we're fielding new technologies and we're not really sure how they're going to interact in a crisis.

So maybe we should pace ourselves. Maybe we should approach international crises more cautiously, and that might also then allow more time to gain the experience you need to develop techniques, procedures, and practices that may mitigate some of these risks that the Roundtable identifies.

The final thing is the human factor. We talk about this in the introduction as well, that one of the things that we've seen from the past—I gave the examples

of at least two incidents in which humans intervene to prevent accidental escalation to nuclear weapons use—is keeping human beings meaningfully in the loop.

We don't quite know what that means right now, because obviously, human beings interacting with systems such as ones that include artificial intelligence, analysis, integration, and decision-making aids are really sort of in a system with each other, and the human's willingness to trust the system varies over time, as Mike Horowitz points out.

But what he does show at the end of that whole timeline, again, is that humans start to correctly assess what the impact of the technology is and to mitigate for the risk and come up with best employment practices. So I think with time and some humility and caution, which maybe is not a common trait among our current leaders, I think we can get there just as we once got there with nuclear weapons.

Sheena Chestnut Greitens: So we were heading toward a note of optimism, and I'm almost afraid to ask this follow-up question because I might tank us again, but I do want to ask you this because your very last phrase there raised this question that kind of jumps out at me when I think about the world today.

And so if you're talking about keeping humans in the loop and relying on proven systems rather than fully automating decision-making in the nuclear realm or the realm of emerging technologies and strategic stability more broadly, given the bureaucratic factors, the organizational cultural factors, and the competitive pressures that exist among different great powers today. When I think about the decision-making systems and processes involved—whether it's contemporary China, or Russia under Putin, or the United States—how realistic do we think that is, not just in the United States, but globally, when we look at the leaders of the countries where this is going to be a key question for strategic stability?

Harold Trinkunas: It's uncertain. This is new territory, really, for the three countries involved. I think the degree of personalism that's emerged in the United States, and in China, and in Russia, is different, in kind, from the systems we once knew, and the systems that were designed to handle nuclear weapons were designed to be part of maybe a bureaucracy that emphasized maybe rationality and strategic calculation.

So I think you're right. This is a pessimistic note, that [we're] heading for a world in which there's more personalist regimes, and where leaders of

personalist regimes that don't already have nuclear weapons are seeking to acquire them is a more dangerous world. Maybe that's a pessimistic note for this conversation, but it's also, I think, a reality.

Sheena Chestnut Greitens: It seems like it is an area where great caution and maybe great concern is warranted. But that's part of why I think the Roundtable is such an important conversation, and each of the pieces in it has something really valuable to say, whether it's to highlight risks or offer some positive steps that might mitigate or reduce some of the risks and the uncertainties.

So, that's one of the reasons why we're excited, and, you know, I appreciate you coming on and doing this podcast today to get us started. In the weeks ahead, we'll be doing podcasts with each of the authors in the special issue and talking in more depth with them about what their essay, their article, has to say about the role of emerging technologies, strategic stability, and the way that human psychology and biases enter into that.

So I'm really excited for this set of upcoming conversations, but [I] really want to thank you and Herb for putting this special issue together. It's been a pleasure to work with you on [this], and we're really excited to be able to share it both in written form at TNSR.org and then also, you know, in this podcast in the episodes that are coming up. So for anybody listening, feel free to tune in, and we'll go through the issues with each of the authors.

Harold Trinkunas: Thank you very much for giving me this opportunity to talk about this great set of papers. Herb and I were thrilled by how it all worked out, and we thank *TNSR* for taking a chance on us. We're really looking forward to seeing both the articles in print and, of course, the podcasts.

Sheena Chestnut Greitens: Thanks for joining us on *Horns of the Dilemma*, the podcast of the *Texas National Security Review*. Our guest today has been Harold Trinkunas, author of the article, "Emerging Technologies and the Future of Strategic Stability," which, as always, can be accessed for free on our website with the rest of the special issue at TNSR.org.

If you enjoyed this episode, be sure to subscribe and leave a review wherever you get your podcasts, and you can always find more of our work at TNSR.org. Today's episode was produced by *TNSR* Digital and Technical Manager Jordan Morning and made possible with support from The University of Texas System. This is Sheena Chestnut Greitens and Ryan Vest. Thanks for listening.