Metis

Study

Deterrence in the 21st century

No. 16 | May 2020

The views expressed in Metis Studies are those of the authors. They do not reflect the opinion of the Bundeswehr, the Federal Ministry of Defence, or the Bundeswehr University Munich. The primary target audience of Metis Studies are practitioners. Metis Studies are based on analyses of scholarly literature, reports, press articles and expert interviews with academics, think tank analysts and policy-makers. References are omitted. Inquiries about sources can be directed at the author(s) via email.

Institute for
Strategy & Foresight

Summary

At first sight, deterrence reflects a simple idea: A credible threat of retaliation convinces potential attackers that the cost of an act of aggression will outweigh its benefit. That is how deterrence prevents war. A closer look at the concept, however, has raised questions ever since the beginning of the nuclear age. Against the backdrop of current security policy challenges, this study examines if and how the concept remains applicable.

The concept of deterrence

The concept of deterrence can be traced through the history of political and military conflict. However, it was not until the Cold War – when it took on the form of nuclear deterrence – that the concept gained the academic and practical prominence it has to this day. Deterrence can work in two ways: by threatening retaliation, i.e. deterrence by punishment, or by denying success, i.e. deterrence by denial. The former involves state A aiming to convincingly signal state B that a certain action or attack will cause prompt retaliation by state A, including the destruction of essential assets in state B. The latter sees state A demonstrating resistance against state B without threatening retaliation, thus suggesting that state B will not be able to achieve its political and military objectives through attack. The two forms may overlap.

Deterrence has often been pursued in aid of strategic ambiguity. For example, during the Cold War era, strategic ambiguity was the preferred approach of the US when it came to Germany. It was impossible to defend isolated West Berlin. Its protection was thus ensured through the threat of retaliatory measures elsewhere. The Soviet Union could never be sure of the place and scale of such measures. The response from Moscow was to exercise restraint.

This example also illustrates the concept of "extended deterrence", which seeks to prevent attacks against third parties such as allies or partner states. Examples include the US nuclear umbrella over Europe or the US security guarantees in Asia. "Direct deterrence", on the other hand, serves only to prevent attacks against a country's own territory. Mutual deterrence between two actors who are both convinced that they are able to destroy their opponent in a retaliatory strike (i.e. mutual assured destruction) establishes a state of strategic stability.

Deterrence does not just happen on its own. The "delicate balance" must be established by political and military means. Its effect is systemic, i.e. it has an impact on international relations. Most important for its effectiveness, however, is its influence on (individual) decision makers.

Theory and practice of deterrence

The academic and theoretical examination of the concept of deterrence emerged as a first wave of research after World War II because of the need for a political response to the nuclear age. The core concepts as outlined above were developed at that time and clearly influenced by the bipolar order of the Cold War. The first wave emphasised the fear aspect of deterrence in that it literally focused on frightening opponents.

The second research wave in the 1950s and 60s discussed the concept in less emotionally charged terms. References to fear were dropped and replaced with discussions of rational actors, cost-benefit calculations and modelling based on game theory, all in an attempt to infer general conclusions about nuclear strategies. The current mainstream of deterrence theory, explicitly understood as the manipulation of an adversary's cost-benefit analysis, remains rooted in these influential efforts. The key concept of "escalation dominance" – i.e. the ability to always go one decisive, ultimately deterrent step further – also dates back to this period.

The third wave, which emerged in the 1970s, called on cognitive psychology and used case studies to examine whether real-world decision makers act in a truly rational manner. As it turned out, the assumptions formed in the second wave were limited in their validity because, in the real world, the misperceptions, recklessness, ideology and even drug use of decision makers ran counter to the idea of rational deliberation. Moreover, empirical studies revealed that decision makers sometimes sought out conflict for domestic reasons (to retain power, for example) despite the threat of deterrence. All in all, the empirical analysis of deterrence strategies suggested that the concept bears the risk of causing exactly the war it is meant to prevent. In other words, the paradox inherent in deterrence theory – that it requires constant preparation for and the credible threat of that which must actually never happen – had been discovered in practice.

Fig. 1 *Permanent confrontation: Deterrence uses a credible threat of retaliation to prevent military escalation (Checkpoint Charlie, 1961).*

After the end of the Cold War, a fourth wave emerged in response to the decline in interstate wars, the increase in intrastate conflicts and the phenomenon of international terrorism. The focus shifted to asymmetric actor constellations and "rogue states" and, in discussing the motives of suicide attackers and moral values of autocrats, Western concepts of rationality in the mainstream deterrence theory were called into question once more. There has been talk of a current fifth wave, although it is more a hotchpotch of approaches meant to use diplomatic, economic, political and military means to address non-kinetic, cyber-specific, terrorist and hybrid risks as they occur. Instead, it creates even more analytical confusion around the concept of deterrence. With paradoxes and problematic basic assumptions left unaddressed since the Cold War, the concept has long been at risk of being overstretched to such an extent as to render the idea of deterrence meaningless.

regulates the interaction of nuclear weapons states, although how much of it is due to rational deliberation versus sheer fear remains unclear. When it comes to newly emerging security challenges, however, it is obvious that some of the fundamentals required for deterrence to work simply no longer apply.

In the information space, cyberattacks raise the question of whether they may be deterred at all. The key problem when applying deterrence to cyberspace is the so-called attribution problem, i.e. the inability to clearly identify the originator of a cyberattack. When a state is able to locate an attacker, for example in a cybercafé or private home in Asia, that information is only of limited use for identifying the actor who actually carried out the attack. Even if the attack is traced back to a computer centre of a local military force, the risk remains that elements of the cyber architecture of the apparent perpetrator state have actually been compromised by third parties. The issue is made all the more complex by some countries tasking non-state actors with operations in cyberspace. As a result, the apparent perpetrator state maintains plausible deniability, which the attacked state struggles to disprove even with advanced cyber intelligence and time-consuming forensic investigation. A threat of prompt retaliation does nothing in terms of deterrence by punishment if it cannot be directed at anyone in particular.

Deterrence in new domains

The "classical" deterrence theory of the Cold War and its practical application have certainly had some effect, although it may have not been as reliable and generalisable as was hoped during the second wave. It undeniably

Thanks to the efforts of state actors to protect cyberspace and critical infrastructures against cyberattacks, successfully mounting such an attack has become much more difficult in recent years. Adversary network operations have to invest more resources, energy, personnel and time to successfully attack a state's key capacities. Attacking military structures from a laptop in a cybercafé is plausible only in Hollywood movies. We may therefore assume that at least deterrence by denial can be used successfully against many actors with few resources. The concept of resilience plays a major part in this respect, as this study will go on to show.

For one thing, some space scenarios call into question the applicability of classical deterrence concepts. For another, the military use of space creates new scenarios that might affect the "delicate balance" on the ground.

In the Cold War era, limited space capabilities meant there was only ever a very small group of "usual suspects" in orbit. Following advances by private actors, however, we now face an attribution problem when it comes to kinetic effects much like the one for non-kinetic operations as previously described above. Once again, deterrence as commonly practiced has no clear target.

Moreover, new space capabilities such as anti-satellite systems or other weapons[1] may jeopardise, for example, early detection of a nuclear first strike or even the second-strike capability and thus erode strategic stability. In such a case, while the logic of deterrence would remain intact, new risks would emerge which could limit its effectiveness as a stable guarantor for the continued non-use of nuclear weapons. For example, a nuclear weapon state – in a case of "use them or lose them" – might respond to an attack against its space-based capabilities with nuclear retaliation in order to forestall being deprived of its second-strike capability.

Deterrence as resilience

Deterrence has thus accompanied us into the 21st century, with old and new questions raised along the way. A promising, forward-looking approach is to consider deterrence in radically simplified terms of resilience, especially in new domains of application. This would also put to rest the deterrence paradox and other legacies of deterrence theory such as the lingering problem of credibility. Such an approach would use the logic behind deterrence by denial as its starting point.

The underlying idea would still be to convince a potential attacker that their plan is useless because it obviously has no chance of succeeding. But while the classical theory of deterrence by denial would involve signalling in the military context, which bears the risk of being misperceived, deterrence by resilience would simply work as a result of robustness and the demonstrated ability to absorb attacks. It is therefore not a question of defence and subsequent counterattack in response to hybrid attacks, for example, but rather of establishing "absorption dominance" and thus the ability to control how damage unfolds and how long it takes to return to the previous status quo after an attack. For example, if state A remains largely unharmed and unimpressed (i.e. unaffected by major cost) by a hybrid intervention of state B involving fake news or cyberattacks, the probability of a second, similar attack will be drastically reduced. To implement this approach, states would have to focus on developing an all-state capacity for resistance and absorption which comprises all critical areas in order to become more resilient to attacks and disruption.

Such scenarios are not without analytical and practical pitfalls because, as in the above example, an attack may barely be registered as such and if it is, it may be difficult to attribute to a certain actor. The attribution problem thus persists, although it is somewhat less relevant because a strategy of resilience does not imply a threat of retaliation and the lack of attribution would thus not undermine its credibility. The identity and intention of an attacker would initially be irrelevant – as long as they are thoroughly discouraged and deterred from renewed attempts. In other words, a strategy of resilience is itself more resilient than a deterrence strategy, at least when it comes to cyberspace.

The concept of deterrence, especially its nuclear dimension, entails risks but remains an integral part of strategies throughout the world. Not every new aspect of security policy can – or should – be addressed with deterrence. But deterrence theory and the scientific observation of its practical use may stimulate new strategic considerations that reflect the complexity of today's security challenges.

[1] See "Space Security", Metis Study No. 13 (August 2019).

Metis

IMPRINT

Publisher
Metis Institute
for Strategy and Foresight
Bundeswehr University Munich
metis.unibw.de

Author
Dr. Konstantinos Tsetsos
metis@unibw.de

Creative Director
Christoph Ph. Nick, M.A.
c-studios.net

Images
Cover: Graphic by ndul on 123RF
p. 4/5: Photo by USAMHI from
wikimedia

Original title
Abschreckung im 21. Jahrhundert

Translation
Federal Office of Languages

ISSN-2627–0609