
Survival
Global Politics and Strategy

Routledge
Taylor & Francis Group

ISSN: 0039-6338 (Print) 1468-2699 (Online) Journal homepage: http://www.tandfonline.com/loi/tsur20

# Artificial Intelligence: A Revolution in Strategic Affairs?

Kenneth Payne

**To cite this article:** Kenneth Payne (2018) Artificial Intelligence: A Revolution in Strategic Affairs?, Survival, 60:5, 7-32, DOI: 10.1080/00396338.2018.1518374

**To link to this article:** https://doi.org/10.1080/00396338.2018.1518374

Published online: 18 Sep 2018.

# Artificial Intelligence: A Revolution in Strategic Affairs?

**Kenneth Payne**

There are some intriguing parallels between the modern development of artificial intelligence (AI) systems, particularly those utilising 'deep learning' methods, and the development of nuclear arsenals and strategy by the Cold War superpowers. Nuclear weapons were retrospectively described as an 'offset' strategy, a way of compensating for relative disadvantages in conventional weaponry, notably by the US against the USSR. Now, AI is heralded in the US as a 'third offset', following the 1970s development of information technologies as a 'second offset' strategy. The analogy clearly appeals to some in the Pentagon. The comparison is meaningful insofar as both technologies offer new military capabilities that have the potential to transform strategy and to dramatically affect the balance of power. The strategic-studies literature on AI is sparse, which is one important reason to consider salient parallels. But there are some important differences between AI and nuclear weapons, which themselves are illuminating about the future of AI in warfare. In particular, nuclear weapons were arguably less revolutionary than AI, in that they did not alter the psychological essence of strategic affairs. In any case, making the comparison yields some useful strategic insights about AI while avoiding the hyperbole that often distorts the issue.

---

**Kenneth Payne** is a Senior Lecturer in the School of Security Studies at King's College London. He is the author of three books on strategy and psychology, most recently *Strategy, Evolution and War: From Apes to Artificial Intelligence* (Georgetown University Press, 2018).

**Survival** | vol. 60 no. 5 | October–November 2018 | pp. 7–32 DOI 10.1080/10.1080/00396338.2018.1518374

Nuclear weapons would be used by human decision-makers and would therefore be subject to human psychology, even if some theorists sought to use them in an abstractly rational manner. AI systems, by contrast, do not make decisions in the same way as humans would, even if their algorithms are often loosely modelled on human cognition. Additionally, AI will be more useable across a full spectrum of force, and indeed more broadly than force itself, shaping all conflicts from the lowest to highest intensity and the smallest to largest scale.

## Current and future capabilities in AI

Recent gains in sensing, categorising and decision-making by modern AI have been marked and rapid.[1] Much progress has been made by AI systems utilising 'deep learning' methods that model, albeit with considerable abstraction and simplification, the neural processes of human brains. At the cutting edge of AI research, algorithms are demonstrating an increasing capacity to learn without supervision, with limited data for training, and to cope with ambiguous and asymmetric information. These developments move AI beyond rapid, brute-force calculation and pattern recognition within tremendous volumes of data – traditional strengths of machine intelligence. AIs can now win at Go (a game with far too many permutations to be susceptible to brute force) and poker (with its asymmetric knowledge of the stakes).[2] Most ongoing AI research is in sectors that are not explicitly defence related – for example in healthcare, internet search or civilian-vehicle manufacture.

Yet the broad capabilities under development – the ability to flexibly categorise information and use this as a basis for decision-making – have clear strategic utility. Early studies indicate that AI will be useful at both the tactical and the strategic levels.[3] For strategic-level decision-making, however, there is likely to be more ambiguity about the data on which to make decisions and the reward functions that any machine seeks to satisfy. AI's immediate role, it seems, will be in the tactical domain; but that will nonetheless have important strategic implications.

AI is already a military reality. For example, weapons-guidance systems make decisions independently of human input (other than in setting the broad parameters for engagement before any action), and intelligence agencies can use algorithms to identify patterns in large datasets. More dramatic changes are under development. In particular, AI systems will allow autonomous decision-making by networked computer agents, enabling extremely rapid sequential action, even in uncertain operating environments. Intelligent machines adept at making inferences on the basis of confused data will soon be on the battlefield. These agents will learn by studying earlier actions, or by observing the parallel actions of other agents in their network.[4] Deep-learning algorithms can do this already, and some have met complex control challenges, such as driving cars and flying helicopters from first principles.[5] New techniques may be even more efficient, incorporating the ability to learn concepts and relationships from smaller samples than those involved in the laborious training processes that typify deep learning today.

Soon, autonomous and intelligent platforms will be able to manoeuvre faster and employ force with more precision than those operated by humans. Already, an AI system can outperform an experienced military pilot in simulated air-to-air combat.[6] Autonomy and networking, along with other technologies, including nano, stealth and bio, will offer sophisticated tactical war-fighting capabilities on land and sea, and in the air. Consider, for example, the impact that shoals made up of autonomous underwater robots sensitive to tiny distortions in the earth's magnetic field would have on traditional submarines. Deployed around bottlenecks in the ocean, such platforms could complicate efforts to conceal the submarines, which currently provide an assured second-strike capability to nuclear powers. In that way and others, a tactical platform has a strategic impact. In addition to these battlefield roles, AI will transform other military activities, including logistics, intelligence and surveillance, and even weapons design. Collectively, these activities, mostly tactical in nature, will have a transformative effect on the strategy of those states employing them. This is because militaries that can successfully develop and utilise them will experience a dramatic increase in fighting power relative to those that cannot.

Beyond the strategic effect of accumulated tactical victories, AI will also shape strategy by offering insights to elite decision-makers based on the processing of vast datasets. Strategic-level AI operating as an 'oracle' for decision-makers will be able to test accepted wisdom, discarding spurious associations, rejecting pet theories and identifying key vulnerabilities in enemies. Considerable shortcomings in this oracle role will persist into the long term, but these will be at least partially offset by some important advantages. Strategic AI will not be subject to the myriad individual and collective psychological processes at work in human decision-making, including groupthink, confirmation bias, bureaucratic politics, excessive optimism and poor risk judgement.

At present, many pertinent AI technologies are immature. Modern unmanned aircraft in service can operate autonomously, but cannot yet execute the sorts of complex missions that manned equivalents can achieve. Land robots are clumsy on uneven terrain.[7] Sceptics rightly point to previous bursts of enthusiasm for AI, followed invariably by disappointment and stagnation as concepts fail to deliver significant breakthroughs in autonomous decision-making. There is considerable wariness that the hype and publicity surrounding deep learning will not pan out as dramatic breakthroughs in cognition that might approach human-level capacity – for example in satisficing between conflicting goals, or in using imagination and memory flexibly to cope with novel scenarios. The AI of today is rather narrow and brittle – adept in its area of expertise, but not at shifting to new tasks. Nevertheless, the rapid progress in AI research, especially of hybrid approaches that utilise multiple AI techniques, along with increasingly powerful hardware on which to run algorithms, suggests the potential for AI to significantly affect existing military activities in the short to medium term, even if it falls short of simulating human-level cognition any time soon.

An important distinction is the extent to which an intelligent agent is either directed or autonomous at the motivational level. This is a high-tech manifestation of the familiar agent–principal dilemma. An AI whose activities are fully determined by human principals and which faithfully anticipates and delivers on their intentions is less unsettling than one that produces unanticipated outcomes – either because the principal's intention was poorly specified or because the agent developed and followed its own subordinate motivations in an effort to fulfil them. Modern AI faithfully seeks to maximise its exogenous reward function – there is no credible, imminent prospect of a malign, Terminator- or HAL-like machine. But game-playing AIs make radical, unexpected and altogether inhuman moves, and military AI may do likewise. The interior logic of an artificial neural network is presently something of a black box.

One last factor to consider is that the distinction between human and artificial intelligence may itself be blurred by advances in human–AI interfaces, or in artificial biological systems. These hybrid approaches, still rather immature but developing quickly, may in time challenge our existing conception of human intelligence.

## Technology and revolutions in strategy

Claims of revolutionary new technologies reshaping strategy are familiar to strategic-studies scholars. There is, however, limited agreement over what constitutes a revolution. A macro view like Alvin Toffler's posits only three large revolutions – agriculture, industry and information.[8] An even more macro view, encompassing two revolutions, may be more apt for considering AI. The first revolution separates *Homo sapiens* from other primates, via a cognitive explosion some 100,000 years ago that brought about rich social interaction, language, the capacity for self-reflection and empathy with others, and the ability to make tools. These are the foundations of human strategy. A second revolution, now under way, is moving strategy beyond purely biological, human intelligence.[9]

More granular accounts offer a plethora of stimuli for military revolutions, including particular items of equipment, modes of manufacture or concepts for employment. For example, anthropologists talk of cultural 'packages'. The horse package entailed the domestication of horses, as well as the development of sophisticated technologies – spoked-wheeled chariots, bits, stirrups and so forth. The 'Asian war' package centred around the composite bow. Gunpowder, steam, rifles, tanks and many other inventions have changed the particular character of war. But their effect on strategy more broadly might be considered modest, insofar as they have not affected higher-level principles such as the utility of surprise or the advantages of concentrating force in space and time.[10] Much writing focuses on the technologies themselves, although more sophisticated analyses reflect on their relationship to the societies that have generated and employed them. In the most compelling accounts, cultures are constantly evolving and interacting, while technologies both reflect and transform their societies, including the military.[11]

AI, like previous military technologies, is a cultural artefact, in this case reflecting the particular structure of economically advanced and largely liberal societies, capable of attracting talent and speculative capital, and robustly protecting intellectual-property rights and academic freedoms. Yet AI is qualitatively different. If earlier technologies transformed the character of conflict and the societies waging it, they left intact its essentially psychological essence – that is, the prosecution of strategy by evolved, embodied and encultured human minds. Insofar as AI departs from this, it may warrant the tag 'revolutionary'.

### *A nuclear revolution?*

Bernard Brodie, a leading interpreter of Carl von Clausewitz and an ostensible exponent of the enduring principles of strategy, was apparently moved on hearing about the atomic bomb to declare that everything he had hitherto written was obsolete.[12] It was not. Nuclear strategy *seemed* esoteric because of the sheer scale of possible destruction, which certainly introduced some novelty to thinking about what to target and when. Yet at its heart, nuclear conflict was still classically strategic – an attempt to use violence instrumentally in difficult and uncertain circumstances, with serious consequences. The modalities and calculus of violence had changed, but its psychological essence remained.

Indeed, nuclear strategy was acutely concerned with the psychology of violence – especially of threat and counter-threat, and the possibility of escalation. There were grave concerns about miscommunication and misperception, and the effects of human error. But all these were staples of strategy more broadly. Nuclear weaponry required new thinking, but so too had other, less destructive weapons systems. Robert Jervis later distinguished between 'conventionalists' who held that nuclear weapons had changed little, and those, like himself, who thought them truly revolutionary.[13] Conventionalists saw a ladder of escalation and – grim logic this – some suggested that even if the summit of the ladder were attained with an all-out exchange, it would still be possible to determine a victor: the side that emerged with an improved relative power position.[14]

Jervis demurred. Nuclear weapons, he argued, worked by threatening the *punishment* of aggressors, rather than *denying* them their objectives along the lines of conventional defences. While punishment via retaliation was not itself new in war per se, the sheer scale of any nuclear reprisal meant that both sides were uniquely vulnerable. War would have catastrophic consequences for both sides, not work as a zero-sum game, with one relative winner. For Jervis, the possession of secure thermonuclear 'second strike' (that is, retaliatory) arsenals on both sides meant that it was impossible *ex ante* to know whether limited wars would remain so. The losing side in such limited conflict, for the first time in human history, could essentially annihilate the other at any moment, such that both could be obliterated in a great spasm of sudden escalation. That was unprecedented, and demanded wholly new ideas about statecraft. This revolutionary development, Jervis averred, made war between nuclear powers extremely unlikely. Mutual assured destruction (MAD) imbued nuclear deterrence with a degree of stability. Neither could risk even minor provocations.

Jervis was surely right that nuclear weapons greatly enhanced the potency of retaliatory punishment, but whether that in itself amounts to a military revolution is moot. Even with the vast destructive power on hand, conventionalists were right that denial remained a part of conflict, nuclear or otherwise. Nuclear powers could harden and conceal targets to make them more resilient, blunting attack. And nuclear confrontation, even actual hostilities, might not necessarily result in MAD, because the prospect of intimidating an adversary into capitulation via the threat of escalation was a feature of nuclear war as much as conventional. There was a macabre logic to the arguments of Herman Kahn that 'escalation dominance', where one side threatened to jump up the ladder of escalation, could produce a 'winner' even after thermonuclear exchange.

And just as denial featured in the post-nuclear age, so too punishment was possible in the pre-nuclear era. Not all conventional defences relied on denial: indeed, pre-emptive war turned on getting one's retaliation in first in order to maximise its impact. Nuclear weapons had certainly changed the relative importance of punishment, requiring new thinking, but the essentials of strategy remained. Most importantly, those essentials were inherently psychological.

In the 1950s and 1960s, some theorising about nuclear weapons sought to emphasise 'rational' strategies that might allow force, including nuclear force, to be used in a calculating, deliberate fashion so as to limit the potential for psychological overreactions and miscalculations to prompt nuclear hostilities. This preoccupation led to game-theoretical modelling of likely strategic behaviours, and reflected a wider trend in the social sciences to borrow from mathematics and the hard sciences. There were benefits to more systematic thinking about nuclear weaponry, including more robust arms control and the elimination of acute vulnerabilities. But the effort was broadly unconvincing, and when nuclear crises came, the behaviour of the actors – while deliberate and calculating, as John F. Kennedy's was in the Cuban Missile Crisis – bore little relation to abstract rationality.

'Rational' strategy depended on the principal protagonists fully understanding the stakes and being able to control action – both debatable propositions. When they did not anticipate correctly, as with Nikita Khrushchev in Cuba, there might be a crisis. There was also scope, as with the *Able Archer* scare of 1983, for inadvertent signalling to prompt alarm – another route to miscalculation.[15] And always there was the possibility that subordinate agents in immediate control of nuclear weapons might not accurately reflect the intentions of their principals.

On the whole, if nuclear weapons had a revolutionary effect on strategy, it was to limit provocations above a certain, albeit hazily defined, level. In practice, the sobering contemplation of nuclear Armageddon seemed to dampen adventurism among senior policymakers on all sides. But the robustness of the deterrent was always a matter of probability rather than a certainty;[16] and the strategic machinations of opposing powers remained a matter for human psychology, regardless of nuclear science. There were enough close calls in the Cold War to suggest that even if nuclear war was ultimately unwinnable, it was still difficult for statesmen to abjure the timeworn staples of strategy – brinksmanship and the search for credibility included.

Jervis understood that the logic of deterrence via punishment was, in reality, anything but abstractly 'rational'. Indeed, he is a leading theorist of the psychological element in policymaking, and psychology features prominently in his writing on nuclear weapons.[17] It was the very uncertainty and imperfections of policymaking that made it impossible to know what would happen in direct superpower conflict. Like Jervis, Thomas Schelling, perhaps the leading strategic thinker of the Cold War, also saw that the psychological essence of strategy remained intact, such that there would be an element of uncertainty no matter the attempts to impose rigorous rational analysis on the problem.[18]

## AI vs nuclear – parallels and differences

Even if claims of a revolution wrought by atomic weapons are overblown, exploring the analogy offers useful insights for those currently thinking through the strategic implications of artificial intelligence. There are sufficient points of similarity to suggest themes in the future development of AI strategy. Firstly, nuclear weapons and AI are both highly technical scientific developments, requiring coordinated expertise. Secondly, the 'revolution' is concentrated in a few states, and the research involves a degree of secrecy which, coupled with the inherent technicalities, constrains public debate. Thirdly, there are valid ethical and legal concerns about proportionality, discrimination and control of weapons employing the technologies. Fourthly, both technologies have the potential to rapidly transform strategy, the institutions charged with applying it, and society more broadly. Lastly, both have potentially apocalyptic consequences and have aroused intense philosophical debate.

### *Expertise*

Both technologies, nuclear and AI, rest on complex, rapidly developing science, with corresponding degrees of uncertainty and risk. Both are the product of large research and development efforts, sustained over decades, often at financial loss. Both are at the cutting edge of scientific research in their respective eras, attracting star researchers and considerable capital, much of it speculative. And both have obvious military implications – requiring parallel expertise in military organisation and strategy. Accordingly, both demand a dialogue between specialist and generalist communities, including the military, scientific researchers and strategic theorists in the academy, the private sector (especially speculative finance) and politics. Few if any individuals could claim the expertise to speak convincingly across all these domains.

With AI, in contrast to nuclear weapons, there is no large-scale, coordinated and covert attempt to achieve a particular end, but rather a more disparate and eclectic research programme. States are increasingly interested in the security dimension of AI, but much of the research is conducted by various private-sector and university-based entities, and in fields less obviously related to the use of military force. The researchers are often keen to promote their latest achievement, or subject it to peer-reviewed publication, in part to attract funding, demonstrate market value in their applications or promote their brand. As a consequence, public debate on AI implications is considerable, and the interested generalist can become reasonably well informed about the state of the art, if not the engineering complexity involved.

While nuclear technologies were dual-use, the enrichment needed for 'weapons-grade' fuel and the need to develop specialised delivery systems, warheads, missiles and submarines all allowed something of a distinction between military and civilian research. In AI, by contrast, there is no comparable distinction, at least insofar as AI research relates to decision-making under complexity and ambiguity: the very basis of tactical and strategic action. Unfortunately for its well-intentioned designers, AI lends itself to a broad range of civil and military applications, and the desire of scientific experts to sequester themselves from military activities is untenable because their work is generalisable and the epistemic community is physically fragmented, notwithstanding efforts to coordinate resistance to the weaponisation of AI.

Of course, some nuclear theorists wrestled with their consciences, and a few became ardent anti-nuclear campaigners, particularly once the scale of devastation from thermonuclear weapons became readily apparent. But once the theoretical physics had been established and practically demonstrated, what remained were engineering problems, and there were enough willing engineers to address them in the context of wartime duty and Cold War tensions.[19] Then as now, the capacity of scientists to shape the ethics of weapons or constrain their development is limited.

*Ethics*

There is a clear ethical dimension to both technologies, evident in the anguishing dilemmas facing their respective designers. For nuclear weapons, this relates most prominently to the scale of destruction wrought relative to any possible *casus belli*. There is also the unique disproportion between the outcome and the relatively small input required – one bomber obliterating Hiroshima when many hundreds had been required for the fire-bombing of Tokyo. Nuclear weapons can, in fact, be made low-yield, and can be used in sparsely populated areas. And there is no a priori reason to ethically distinguish between a nuclear strike and a raid that requires far greater conventional input but delivers the same explosive power. Indeed, the nuclear option may put fewer of its perpetrator's own forces at risk.

Nevertheless, a strong normative taboo against nuclear weapons has emerged since their first use.[20] Crossing the nuclear threshold in a conflict would now constitute a staggering historic decision, regardless of the size of the bomb or the nature of the target. Meanwhile, a paradoxical situation has emerged whereby the robustness of the deterrent, and hence the prospects for peace in a nuclear crisis, depends on the credibility of an assuredly unethical threat to obliterate large numbers of civilians. Meanwhile, efforts to limit the number and types of weapons held by a state, or to develop effective countermeasures that might mitigate the threat, are potentially dangerous insofar as they challenge the stability of an assured retaliatory strike – thereby encouraging a state to cross the nuclear threshold while it still has a chance. The ethics of nuclear weapons have always had a peculiarly and perversely paradoxical quality.

With AI, meanwhile, ethical concern has centred on the issues of control and accountability.[21] In theory, AI systems will be able to distinguish between combatant and non-combatant, according to parameters established by human handlers. And they will then be able to apply force with great precision, allowing for fine discrimination. Even so, the scope for ambiguity and error is inevitable, as it is for humans operating in the same environment. Moreover, like humans, a machine in combat would have to make difficult philosophical choices about the value of life: should it adopt a consequentialist logic, in pursuit of the greater good, and if so, from whose perspective? Or does it have a duty to each individual life? The just-war tradition, whose precepts the machine would likely be following, is a blend of consequentialist and deontological ethics, making consistent application impossible; and humans frequently fudge the issue in the moment, reaching judgements about appropriate norms retrospectively. But in specifying rules of engagement for an AI system beforehand, such fudging would be impossible, and no 'expert system' specification would be able to account for all possible contingencies.

There is also a control problem at the *ad bellum* level – that is, when deciding to initiate hostilities. The automation of decision-making in response to fast-moving threats entails a risk of inadvertently initiating hostilities – particularly dangerous if there is an automatic capacity for escalation in response to perceived threats. AIs in the financial sector have already prompted similar alarm, as when automatic trading algorithms have produced drastic market fluctuations. Both *ad bellum* and *in bello*, the speed of AI decisions threatens the human capacity to control strategic events, yet the security dilemma and the evident tactical advantages of rapid automated decisions make keeping a 'man in the loop', or even 'on the loop', problematic.

Fears about loss of control typically focus on unintentional or unforeseen eventualities that confound efforts at *ex ante* specification of the task. But there are other related questions about how to reconcile conflicting goals, cope with ambiguous or subjective goals, or respond to changing goals. And there is the spectre, albeit distant, of an artificial general intelligence (AGI) with incompatible goals of its own. These are all particular manifestations of the agent–principal problem which inevitably arises in group endeavours, including the formulation of strategy. But they are lent a particularly acute dimension by the alien and novel nature of AI, notably its speed and its resolve. AIs don't readily fudge, satisfice or understand the notion of 'good enough'.

*Arms control*

One familiar aspect of the ethical debate is control by states and by international regimes. Nuclear weapons are prohibited by international treaty, and nuclear powers have undertaken to limit practical testing and not to develop new types of warheads. From the outset, control of nuclear weapons was retained by states, despite early proposals to place them under international supervision. At the time, the United Nations was new and untested, and the failures of the League of Nations abundantly evident. Moreover, the Cold War was rapidly intensifying, and the nuclear taboo had not yet been robustly established. The US, for a brief period, enjoyed a monopoly on the technology. The possibilities for defection from any international regime were real, with potentially devastating consequences. Monitoring and restrictions on the very large industrial activities involved in developing and testing weapons later proved feasible, and played a part in stabilising the strategic balance and limiting proliferation. But initially there was great uncertainty about the capabilities of adversaries (manifest, for example, in the suspected 'missile gap' between the United States and the Soviet Union in the late 1950s and early 1960s), and the strategies that they would employ in conflict.

The idea of arms control for AI remains in its infancy. The International Committee of the Red Cross has published advisory guidance on the use of autonomous weapons, but customary and formal international law remains in flux.[22] As with the early days of nuclear weapons, norms will likely follow technology, with law materialising still later. This is because the advantages of possessing weaponised AI are likely to be profound, and because there will be persistent problems in defining technologies and monitoring compliance.

*Armed forces and society*

Alongside its potential to transform the relative power of existing states, AI will effect a transformation of human society itself. For example, dramatic productivity gains from AI in economic activity may favour some economies more than others, which could conceivably produce pressure for federated coordination above the level of existing states to control corporate power in AI research and to manage a huge overcapacity of human labour. AI may also extend the scope for state control of citizens. Particularly in democratic societies, there will be concerns about the extent to which AI undermines prevailing norms of privacy. There will be pressure to amass and analyse vast datasets to identify potential threats to the state, from both indigenous and external actors, broadening processes already under way. Such developments raise standing concerns about representative government and civil liberties.

The extent to which these capabilities will reshape military organisations is also unclear. There will be additional requirements for technical expertise, systems management and integration, but these may be feasible within existing service structures focused on particular domains. Tactical air, where linked to the activities of ground forces, could logically be delivered by the army. But logic is not the only driver of organisational change. Meanwhile, the proliferation of autonomous unmanned systems will certainly challenge existing notions of the warrior – continuing a trend already under way with drone pilots. It's also probable that fielded systems will include AI and human operators working together, including in using exoskeletons, semi-autonomous tactical platforms and perhaps even forms of brain – and body – machine interface. AI systems will inevitably depopulate the battlefields, and thereby challenge existing sub-unit structures and flatten military hierarchies of medieval origin, rooted in the control and motivation of large bodies of men. New relationships between society in general and the traditional institutions charged with protecting it will likely emerge. Overall, the societal impact of AI will be more profound than the one that came with the nuclear revolution, reaching far beyond military issues.

### *Strategy*

While there are some striking parallels, the differences between AI and nuclear weapons come into sharpest focus when it comes to strategy itself. Though perhaps not truly revolutionary, nuclear weapons nonetheless required the prompt and creative integration of the new weapons systems into existing approaches to strategy. Their development produced a profound break with the past, reshaping the distribution of power, changing the character of warfare and greatly enhancing the destructive force available to states that possessed them. This required the development of new strategic thinking, new organisational structures and new equipment.

From the outset, there was intense debate and deliberation on how best to employ nuclear weapons to coerce or deter adversaries. The key strategic ideas that emerged included the importance of retaining viable second-strike capabilities via concealment and hardening; the advisability of using many bombs (later multi-warhead missiles) to ensure delivery in the face of countermeasures; the tension between counterforce (targeting enemy military capabilities, especially nuclear weapons) and counter-value (targeting civilian populations, thereby holding them hostage to the idea of further nuclear escalation) in terms of what constituted the best deterrent; and the rational application of calibrated force in a crisis to deter an adversary from further escalation.[23] This non-exhaustive sketch does scant justice to a rich and intricate body of strategic thought, but demonstrates that nuclear weaponry required imaginative and rigorous thinking along different lines than did conventional weaponry. This was precisely because, as Jervis argued, its primary effect was to threaten massive punishment, especially against similarly equipped adversaries.

The effect on military organisations was profound. For the United States, the advent of the nuclear age meant an enhanced role for the Strategic Air Command (SAC). The SAC became the dominant branch of the newly formed US Air Force and, arguably, of the US military overall, particularly during the Eisenhower administration, which privileged nuclear deterrence over expenditures on conventional weaponry in seeking a 'bigger bang for a buck'. The development of nuclear-strike aircraft – heavy bombers and strike fighters alike – dramatically shaped the organisational ethos of the service, its procurement practices and its doctrine. Soon, however, the development of reliable submarine-launched ballistic missiles and stealthy boats allowed the US Navy to gain greater prominence in nuclear deterrence. In the Soviet Union, meanwhile, the balance of effort went into developing land-based fixed and mobile intercontinental-ballistic-missile (ICBM) capabilities that optimally exploited the depth of the Eurasian landmass. Strategic precepts for the employment of nuclear weapons also differed, with no discernible effort to mirror the arcane logic of game-theoretical approaches advanced by many American thinkers.

The ways in which the two nuclear superpowers adopted the same technology were thus heavily bound to their respective military cultures and distinctive strategic circumstances, albeit with some points of commonality. Both societies, democratic and totalitarian, attempted to establish civil-resilience programmes and contemplated anti-missile defences to protect their populations. Both engaged in extensive espionage and reconnaissance programmes to understand the capabilities of their adversaries. Both developed strategies for extended deterrence to reassure allies. Both fought proxy wars at the peripheries of their blocs, and otherwise sought, with some success, to limit direct provocations that might produce escalation to direct hostilities.

Moreover, there was a consensus that nuclear weapons could not be strategically accommodated merely by thinking that a nuclear bomb was just an unprecedentedly powerful offensive weapon that could be introduced into theatre-level crises without having a dramatic impact far more momentous than its TNT equivalent. Thus, nuclear weapons transformed armed forces, and their relationship to wider society, not least by necessitating a powerful and enduring national-security apparatus that was extensively entwined with industrial and academic activity.

These radical changes notwithstanding, there has been considerable continuity in strategic affairs in the nuclear era. Tradition and inertia are important influences on strategy, perhaps because crises are infrequent and dissimilar, limiting the efficiency imperative. The persistence of amphibious and airborne capabilities incommensurate with their utility provides examples, as will the enduring preference of pilots for manned flight. In earlier eras, the preference for mounted cavalry and reluctance to embrace armour reflected comparably resilient traditions. The conservative approach to nuclear weapons by senior military officers contrasted with the more radical rethinking of strategy by civilian specialists, including many drawn from outside the ranks of strategy specialists. But nuclear strategy was conservative at an even more fundamental level than cultural stickiness, insofar as it remained an inherently psychological phenomenon. In Clausewitzean terms, the persistence of friction, uncertainty and chance left decision-making firmly an art more than a science, requiring judgement of risk and the capacity to act on incomplete information.

## AI and strategy

In contrast to strategic thinking about nuclear weapons, that about AI is immature. Among the key considerations are speed and command and control. Then there are concerns unique to the technology: the capacity of AI to cope with ambiguous and rapidly evolving data and to learn from limited data; its ability to intuit complex, associative meaning and develop imaginative responses; and its capacity to effectively interpret and execute the human intentions that underpin its activities, even where these are themselves complex and multifaceted. The uncertainty over the precise capabilities that will emerge and their distribution among states complicates efforts to discern broad strategic principles for AI. Nevertheless, some themes have already emerged. While the strategic fundamentals of air, sea and land power are well understood, and have withstood many technological changes, the ability of AI to seamlessly connect disparate domains and to dominate through the speed and accuracy of its thought, manoeuvre and fire capabilities will challenge some of these long-standing strategic standards, including those that relate to nuclear weapons. What can be said already about these changes?

Firstly, AI will change power balances. AI systems will undoubtedly enhance the ability of militaries that possess them to reconnoitre, manoeuvre and employ deception, before rapidly concentrating force and delivering precision fires. This will change the utility of force by enhancing lethality and reducing risk to societies possessing AI war-fighting systems. Effective AI will likely overmatch legacy military capabilities, dramatically redrawing the balance of power. Moreover, a marginal technological advantage in AI is likely to have a disproportionate effect on the battlefield, given that small advantages in decision-making ability, notably in terms of speed and accuracy, can translate disproportionately into dominance. The key question then becomes who gets what sort of AI, and how quickly? There are some key technical barriers to entry, which suggest a variegated uptake and limited capacity for others to innovate or emulate. This probably favours existing advanced industrial societies such as the US, Europe and perhaps China. These societies will see their military power enhanced relative to others, well beyond the enhancements already realised through the information revolution in military affairs.

The distinction most relevant will be between the best algorithm and the rest. That's because marginal quality might prove totally decisive: other things being equal, we can expect higher-quality AI to comprehensively defeat inferior rivals. In contrast, even rough-and-ready nuclear arsenals, like that possessed by North Korea, can deter more sophisticated adversaries. But while AI quality will count, antebellum uncertainty about whose AI is best will complicate power assessments and may be destabilising. In this respect at least, AI is comparable to earlier conventional weapons technologies.

The route to dominant AI need not be linear. Local optimisations generating a temporary strategic advantage will not necessarily preclude the eventual emergence of better alternative approaches. If the local optimisation is sufficiently advanced, however, it might trump all comers decisively enough to stymie other approaches. Alternatively, a marginal advantage in AI quality may accelerate, perhaps even towards 'superintelligent' AGI capable of more flexible and self-directed learning. Under the most dangerous scenario, one power threatens dominance and induces rivals to court great risk to avoid that outcome. A nuclear deterrent would mitigate this alarming prospect, but only if the adversary's emerging AI capability did not wholly preclude a retaliatory second strike. Another, perhaps more likely, possibility is that uncertainty about the relative capabilities of rival AIs, and thus of the distribution of military power at any moment, could provide scope for considerable brinkmanship, and therefore miscalculation. The current multipolar distribution of power further compounds the difficulty of gauging power reliably when it rests on a fast-changing capability. If legacy military power is rapidly overtaken by

AI, the overall distribution of power might switch from multipolar to unipolar with unprecedented speed.

In assessing the impact of AI on the balance of power and incidence of conflict, much will depend on the speed with which any given AI is acquired, and the extent to which there are significant barriers to proliferation. In broad terms, gains will probably be rapid (given current rates of progress) and unevenly distributed (given technical challenges of emulation and bottlenecks in expertise). A great deal also hinges on the particular qualities of any given AI. It's likely that the more generalisable and flexible the intelligence created, the greater its military advantage. But the first military AIs will be specialists and operate at the tactical level. Narrow, domain-specific AIs may outperform more general-reasoning AIs in their area of expertise for a long while to come, given the considerable challenges of conceptualising and coding AGI.

Secondly, AI changes the risks from using force, especially for casualty-averse states, which are most likely to field it. The net effect is unclear. Changes in the pay-offs associated with fighting may actually provoke conflict by making it affordable for hitherto risk-averse states. Elsewhere, however, AI could deter aggression by adventurers seeking easy gains that are no longer below the threshold for intervention. Alliance relationships could become complicated: alliance members not possessing cutting-edge AI capabilities could be correspondingly reluctant to engage in operations that were still risky for them. Such allies might be useful legitimisers for action, without contributing themselves, reducing them to client-state status, effectively demilitarising states content to function under a Pax AI.

The varied distribution of capabilities and significant barriers to entry echo concerns from the nuclear era about blackmail by possessing states and preventive strikes by states fearing what rivals might do with the technology if they obtain it. Strategically, the options are also familiar for the laggards – balancing against AI-possessing powers (difficult, given the qualitative edge) or bandwagoning alongside them. Geography and culture (including ideology) provide rationales for both options.

Thirdly, in contrast to nuclear weapons, which heavily favour defence provided they can survive a first strike, AI should favour the offence, given its speed, precision, and acquisition and analysis of unbiased (by human heuristics) knowledge. While such attributes can equally be utilised by militaries on the defensive, there are two important respects in which AI shifts the balance. Most obviously, the offence by definition has the initiative, and with mature AI that alone might be sufficient to overwhelm defences. Secondly, Clausewitz's remarks about the culminating point of attack and the relative strength of the defence have an underlying psychological quality that he himself articulated, and that subsequent research has demonstrated: humans are loss averse, place greater value on possessions in hand than those sought, and are prone to gamble more when losing than when gaining.[24] What we have, we hold. AI, by contrast, is not susceptible to these human tendencies. Its resolve and appetite for risk is not shaped by a subjective, psychological anchor favouring the defender.

Among other considerations are nuclear ones. Weak states have been able to effectively deter stronger ones by dint of acquiring a nuclear weapon for its tremendous defensive strengths. But AI enhances the possibilities for successful first strike against adversaries possessing limited nuclear arsenals, and could even shift the balance against adversaries that are better endowed with nuclear weapons.

Fourthly, AI will shape military activities across a full spectrum of violence, rather than preserving the clear normative distinction between conventional and non-conventional systems that evolved in the case of nuclear weapons and the associated taboo. Because its utility ranges from the most minor levels of force to massively destructive thermonuclear power, AI could allow possessing states to achieve escalation dominance against conventionally equipped adversaries at any intensity, and for any type, of hostilities. Against other AI-possessing states, the calculus is unclear. An evident danger would arise in automating escalation, notwithstanding the imperative to act with speed against peer adversaries inclined to do likewise. Insofar as AI's advantages lie in speed and decentralised control, this dynamic is hazardous and demands close attention to the specification of goals by humans.

Further complicating the picture, AI methods apply not just to explosive force, but to any number of military processes – logistics, weapon design and human resources, to name a few. And AI itself is not a discrete, clearly bounded category, like a nuclear weapon. Accordingly, the distinctions involved are unlikely to be binary, as they would be between nuclear and non-nuclear states, or nuclear and non-nuclear explosives. This complicates both strategy and regulation. Some AI technologies may not be applicable across the spectrum of activities, and certain states may enjoy a comparative advantage in some but not others.

### Strategy is psychology

One overarching theme runs through these four strategic areas – the role of human psychology in strategy. Until now, humans have been gauging uncertainty, weighing risk, making judgements about relative power and resolve, and deliberating on questions of escalation and deterrence. But these calculations increasingly need not necessarily be made by humans, or at least humans unaided by AI.

The human biases and cognitive heuristics that had proved adaptive in an evolutionary setting presented a grave threat to humanity when they could be applied to nuclear fusion reactions. But there's no a priori reason to suppose that supplanting them with AI will necessarily be safer. Many human heuristics work to streamline decision-making by reducing cognitive load. This sometimes produces maladaptive decisions, but not inevitably. Strategy as crafted by an AI need not be susceptible to fatigue or emotion, or be influenced by any of the battery of cognitive heuristics that are the signatures of human intelligence. But what it *would* be shaped by is a matter of considerable debate. And, critically, whatever its cognitive processes, its success as a strategic agent depends on its ability to interpret and execute human wishes.

An AI is not an embodied and intensely social animal, and does not have biologically and environmentally evolved emotions and motivations. In time, an AGI may develop its own motivations, or it may be programmed with some facsimile of human motivations or supporting emotions at the outset. None of that is a given – emotions and motivations, many of which remain obscure to the conscious deliberation of those actually experiencing them, are distinctly biological adaptations. It may be that the major advantage of AI, in addition to its tremendous speed and capacity to process great volumes of information, lies in its *lack* of susceptibility to such human traits. In that case, the AI would simply be working to optimise its given task without human cognitive tendencies.

Processing speed and memory in machines accelerate decisions, and artificial heuristics can be employed too. The result can be timely and accurate decision-making, even in information-rich environments. But there are dimensions of complexity that befuddle even the best AI, particularly when it comes to categorising and contextualising information in ways that approximate human meaning. This is partly, though not exclusively, due to the absence of those biological foundations that provide much of the context for our decision-making.

For example, piqued prestige may animate human strategy and motivate those executing it, providing a fillip for fighting power. Human goals and means in strategy often emerge this way, in the moment. Humans often don't know what they want, or how badly they want it, until the shooting is under way. An AI that is not motivated by prestige might not respond similarly to a provocation. To be faithful to humans, that AI would need to be able to respond dynamically to their emerging, revealed preferences. Insofar as human motivations and cognitive heuristics produce sub-optimal decisions, their absence in an AI could be considered a good thing. But insofar as human strategic goals and methods reflect the intentions and desires of the human agents making them, it might not.

So the danger in AI, whether employed for a tactical weapons system or a strategic-scenario planner, lies primarily in the gap between how the AI solves a problem framed by humans, and how those humans would solve it if they possessed the AI's speed, precision and brainpower. This is a particularly striking example of the principal–agent problem common to all social affairs, including strategic matters. Humans have evolved social mechanisms to manage our group relations, even if imperfectly. With AI in the mix, the principal–agent problem is compounded by the speed of devolved AI decision-making and the radical divergence in cognitive processes of human and AI agents.

To meet human expectations, AI will need to faithfully reflect the constantly evolving human perception of conflict. But the question remains whether it should act on what humans wanted at the outset, what they want in the present moment, or what the AI understands to be most closely suited to the future they might want. There is scope for considerable differences among all three – with profound implications for whom the AI targets, and how.

Clausewitz originally thought that conflict would escalate towards some idealised total war, given the passions inherent in it. Later, however, he modified his view, observing that war in the real world was invariably limited – both by the capabilities available to the belligerents and by their desire to pursue particular goals through violence. War involving AIs need not be subject to the same limitations. If the AI miscalculates what its human principals want, there is tremendous potential for catastrophic and unwarranted violence. The challenge extends across all dimensions of strategy, from the most tactical activity upwards. Humans can impose rules of engagement, only to find that in execution they do not match evolving societal goals. An AI that escalates to meet pre-specified criteria of reputation and credibility may not be able to reflect on the consequences of its actions in time to allow a change of course.

When crises arose in the nuclear era in Berlin and Cuba, the policymakers and military elites involved proved themselves possessed of a distinctly human 'rationality', not the abstract game-theoretical one that some strategists advocated. Robert F. Kennedy's insider account of the crisis and the tapes of the ExComm meetings themselves capture the intense, emotional dimension of the strategists' deliberation.[25] The picture emerges of actors feeling their way forward amid great confusion and fear. The lesson here is that human psychology is not necessarily a weakness in adopting effective strategy. Taking the emotion out of strategy via game theory, with its cool calculus of escalation dominance, was not ultimately possible. With AI it will be – and therein lies AI's true radicalism, and perhaps its greatest risks.

* * *

There remains great uncertainty about the implications of AI. On one hand, Stephen Hawking and Elon Musk claim that there is great peril for humanity. On the other, AI experts like Yann LeCun and Geoffrey Hinton contend that AI has made only faltering advances towards AGI, and so remains a limited threat to humanity, whether by itself or as a servant of humans. So far, the incorporation of AI into military systems has been limited. Yet, regardless of the timescale, AI that is adaptive to a range of decision-making contexts does presage a profound shift in strategy that is likely to be even more radical in nature than the nuclear revolution. The apocalyptic view of AI focuses on the possibility of its own intrinsic motivations. The dystopian sci-fi vision, of the sort portrayed in movies such as John Badham's *WarGames* (1983), Alex Garland's *Ex Machina* (2014) and other popular-culture vehicles, centres on an anthropomorphised superintelligence that is motivated to oppose humans. But long before such an eventuality arises, the rapid advances of AI that seeks to optimise human goals is beginning to transform military activity, and demands new strategic thought.

There are many parallels with the onset of the nuclear era – not least the great strategic and scientific uncertainty and the tremendous increase in military power. Still, it is the differences that are most compelling – specifically, in the sorts of cognition that will be applied to strategic issues. AI will not be easily regulated because it is a remarkably heterogeneous and rapidly evolving field, the products of which apply to any form of complex decision-making. And it will be especially tricky to align the execution of AI-enabled strategy with our human intentions. But these essentially technocratic considerations should not obscure the feature of AI that makes it so revolutionary. For the first time since the cognitive revolution began tens of millennia ago, human strategy may be shaped by non-biological intelligence that is neither embodied nor encultured.

## Notes

1 Margaret A. Boden, *AI: Its Nature and Future* (Oxford: Oxford University Press, 2016); David Vernon, *Artificial Cognitive Systems: A Primer* (Cambridge, MA: MIT Press, 2014).

2 See David Silver et al., 'Mastering the Game of Go with Deep Neural Networks and Tree Search', *Nature*,

no. 529, 28 January 2016, pp. 484–9, https://doi.org/10.1038/nature16961; and Nikolai Yakovenko, 'Artificial Intelligence and Hold'em, Part 1: Counter-factual Regret Minimization', *Poker News*, 23 October 2015, https://www.pokernews.com/strategy/artificial-intelligence-hold-em-1-23152.htm.

3 Kareem Ayoub and Kenneth Payne, 'Strategy in the Age of Artificial Intelligence', *Journal of Strategic Studies*, vol. 39, nos 5–6, 2016, pp. 793–819, https://doi.org/10.1080/01402390.2015.1088838. See also Michael Horowitz, 'Artificial Intelligence, International Competition and the Balance of Power', *Texas National Security Review*, 15 May 2018, https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/; and Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W.W. Norton, 2018).

4 Sergey Levine, Timothy Lillicrap and Mrinal Kalakrishnan, 'How Robots Can Acquire New Skills from Their Shared Experience', Google Research Blog, 10 March 2016, https://research.googleblog.com/2016/10/how-robots-can-acquire-new-skills-from.html.

5 Pieter Abbeel, Adam Coates and Andrew Y. Ng, 'Autonomous Helicopter Aerobatics through Apprenticeship Learning', *International Journal of Robotics Research*, vol. 29, no. 13, 2010, pp. 1,608–39, https://doi.org/10.1177/0278364910371999.

6 Nicholas Ernest et al., 'Genetic Fuzzy Based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions', *Journal of Defense Management*, vol. 6, no. 1, 2016, https://doi.org/10.4172/2167-0374.1000144.

7 Matt Simon, 'Stop Laughing at Those Clumsy Humanoid Robots', *Wired*, 11 June 2015, http://www.wired.com/2015/06/stop-laughing-clumsy-humanoid-robots/.

8 Alvin Toffler, *The Third Wave* (London: Pan Books, 1981).

9 See Kenneth Payne, *Strategy, Evolution and War: From Apes to Artificial Intelligence* (Washington DC: Georgetown University Press, 2018).

10 See Beatrice Heuser, T*he Evolution of Strategy: Thinking War from Antiquity to the Present* (Cambridge: Cambridge University Press, 2010).

11 See, for example, Azar Gat, *War in Human Civilization* (Oxford: Oxford University Press, 2006).

12 See Barry Scott Zellen, *State of Doom: Bernard Brodie, the Bomb, and the Birth of the Bipolar World* (London: Continuum, 2012).

13 Robert Jervis, *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon* (Ithaca, NY: Cornell University Press, 1989).

14 See, for example, Herman Kahn, *On Thermonuclear War* (Princeton, NJ: Princeton University Press, 1960).

15 Arnav Manchanda, 'When Truth Is Stranger than Fiction: The Able Archer Incident', *Cold War History*, vol. 9, no. 1, 2009, pp. 111–33, https://doi.org/10.1080/14682740802490315.

16 Michael Quinlan, *Thinking about Nuclear Weapons: Principles, Problems, Prospects* (Oxford: Oxford University Press, 2009).

17 See, for example, Robert Jervis, *Perception and Misperception in*

*International Politics* (Princeton, NJ: Princeton University Press, 1976).

18 Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 2008).

19 Joseph Rotblat, *Scientists in the Quest for Peace: A History of the Pugwash Conferences* (Cambridge, MA: MIT Press, 1972).

20 Nina Tannenwald, 'The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use', *International Organization*, vol. 53, no. 3, Summer 1999, pp. 433–68.

21 Ugo Pagallo, 'Robots of Just War: A Legal Perspective', *Philosophy & Technology*, vol. 24, no. 3, September 2011, pp. 307–23, https://doi.org/10.1007/s13347-011-0024-9.

22 International Committee of the Red Cross, 'Autonomous Weapon Systems: Is It Morally Acceptable for a Machine to Make Life and Death Decisions?', Statement of the International Committee of the Red Cross, 13 April 2015, https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS.

23 Lawrence Freedman, *The Evolution of Nuclear Strategy* (Basingstoke: Palgrave Macmillan, 3rd edition, 2003).

24 Kenneth Payne, 'Prospect Theory and the Defence in Clausewitz's "On War"', paper presented to the annual meeting of the International Studies Association, Atlanta, GA, 16 March 2016, http://web.isanet.org/Web/Conferences/Atlanta%202016/Archive/968b741d-130c-4912-a4ee-997345a57ce1.pdf.

25 See Robert F. Kennedy, *Thirteen Days: A Memoir of the Cuban Missile Crisis*. (New York: W.W. Norton, 1969); and Ernest R. May and Philip Zelikow, *The Kennedy Tapes: Inside the White House during the Cuban Missile Crisis* (Cambridge, MA: Belknap Press, 1997).