Discrete, specified, assigned, and bounded problems: The appropriate areas for AI contributions to national security

Invited Perspective Series

Strategic Multilayer Assessment (SMA)
Future of Great Power Competition & Conflict Effort

DECEMBER 31

STRATEGIC MULTILAYER ASSESSMENT
Author: Dr. Zac Rogers
Series Editor: Sarah Canna, NSI Inc.

This white paper represents the views and opinions of the authors. This white paper does not represent official USG policy or position.

Dr. Zac Rogers

Dr. Zac Rogers PhD is Research Lead at the Jeff Bleich Centre for the US Alliance in Digital Technology, Security, and Governance at Flinders University of South Australia. Research interests combining national security, intelligence, and defense with social cybersecurity, digital anthropology, and democratic resilience.

Discrete, Specified, Assigned, Bounded Problems: The Appropriate Areas for AI Contributions to National Security

Dr. Zac Rogers

The cluster of technologies associated with artificial intelligence (AI) research and development, and the implications of these technologies across the spectrum of human activity including national security, generates a blur of fast-moving technological, political, strategic, and ethical puzzles. For all their hyper-modern veneer, tracing the trajectories of historical thought associated with these puzzles offers some much-needed clarity. To this end, what follows is an overview of three relevant trajectories. While far from exhaustive, practitioners and theorists alike in national security should have a solid grounding in these thought trajectories to facilitate reasoned decisions about AI research and development.

The overall argument asserts that research and development of AI technologies for national security should be confined to areas where discrete, specified, assigned, and bounded problems and tasks can be scientifically explored and assessed. Battlefield AI for intelligence, surveillance, and reconnaissance (ISR), target identification, sensing, and weapons tracking is such an area. The Joint Artificial Intelligence Center's current four focus areas are appropriate (United States Department of Defense, 2019). Areas of indiscrete, unspecified, unassigned, and unbounded problems and tasks, such as in the socio-political realm incorporating population-centric cognitive and information warfare, should be approached with a high degree of caution. Risk-taking in this area is attended by high degrees of uncertainty with high exposure to catastrophic costs to domestic populations. Policy-making in AI for national security should follow a bi-modal strategy which allocates the appropriate cost/risk ratios to maximize adaptive innovation and minimize hubris.

Positivism, Behaviorism, Dualism, and AI: Two Resurgences and Some Sleight-of-Hand

Positivism–A Short History

The onrush of digital information technologies has recapitulated the controversies of Positivism. The core Positivist commitments are to the primacy of empirical science in knowledge-making, to the applicability of that knowledge to every facet of human affairs including the social, economic, and political realms, and to the inevitable convergence of moral, social, economic, and political forms based on that application. Empiricism conceives of knowledge as only (or primarily) composed of the data we receive through sensory experience. Positivism added a normative element; it takes empiricism a step further by asserting true knowledge can be validated empirically and that empirical science is, therefore, the only rational basis for organizing society.

For the Positivist fountainheads Henri de Saint-Simon (1760-1825) and Auguste Comte (1798-1857), this Enlightenment commitment meant all other historical forms and traditions of knowledge-making were certain to be swept away, as a superior scientific modernity was inevitably converging. A singular modernity espousing a singular morality would emerge, in which humankind would use science and technology to overcome resource scarcity, and put an end to poverty and war. Without conflict, there would be no need for power competition.

Marx borrowed from Saint-Simon when he famously declared communist rule would culminate in an 'administration of things.' The French Positivists had significant influence on the 20th century progenitors of Bolshevism, Maoism, and Nazism, all of which in various ways expressed a commitment to the inevitable superiority of the 'scientific society' according to their own interpretations. The appeal of many of the tenets of Positivism, particularly in the recourse to physiologically determined social strata, to the European inter-war Far-Right was undeniable.

Aiming to do more than merely revolutionize society, Saint-Simon and Comte were famous for their development of a secularized 'Religion of Humanity,' endorsed by liberal theorist John Stuart Mill with whom Comte corresponded. Expressing an unlimited faith in the power of social engineering, Comte aimed to complete the Positivist project by turning it into a fully-fledged religion in which Man replaces God. For liberal humanists, Positivism's elevation of human rationality and its promise of manageable melioristic progress was also highly attractive.

Positivism's chief intellectual aim was to develop an entirely physiological account of society and human affairs, an account Comte dubbed 'social physics,' and one we find repeated in the immensely popular contemporary pseudo-science of behavioral economics (1856). For Comte, if society were to be studied like any other physical system, its empirical basis would be the quantification and measurement of human behavior. Here Positivism's commitment to empiricism collides with one of the chief controversies within psychology and the science/philosophy of mind known as Behaviorism. If Comte's new science of 'sociology' were to progress, it must offer a complete account of society and human affairs, which meant it must capture the essential component of human life–the mind. The chief controversy of Behaviorism is thus: to what extent can the quantification of human behavior account for the subjective inner states of the human mind? Does it even need to?

In the 19th century, this French claim to embody universal civilization gave rise to the German counter-movement of Romanticism, expressed in the work of thinkers such as Nietzsche and Fichte, who glorified Will over Reason. The fact, as the 20th century demonstrated, that the envisioned singular modernity did not emerge dealt Positivism a mortal blow. With the proliferation of the tools and methods of science and technology, and the Promethean power they offer humans over their surroundings and other humans, a fragmented, conflictual, and multiple-modernity is unequivocally what emerged by the turn of the 21st century. In what is an underacknowledged blow to the many species of Enlightenment thought, neither reason nor will have proven capable of containing the protean power of technology.

Neo-Positivism and the Digital Age

As the regime of AI sciences increasingly offer the national security, intelligence, and defense (NSID) communities a range of tempting opportunities, Positivism's controversies must again be clarified. Remarkably, Positivist ideals live on in many of the hopes and dreams of the data-driven society. The explosion of empirics, tools, and methods enabled by the digital age have fueled a resurgence in the belief that hidden somewhere in the deluge of data are the hidden codes of human behavior. Ahistorical, acultural, and apolitical, society is a system-of-systems driven by discoverable laws, able to be rendered digitally as behavior, manipulated, and 'nudged' by a new generation of unannounced neo-Positivists wielding advances in AI.

As with the 19th century Romantics, the counter-movement today can be seen in the many global variants of violent extremist organizations (VEO), for whom overcoming the scientific society they associate with the West is an act of will and an anti-positivist—though thoroughly modern—faith in the transformative power of violence. The belief that violence can be wielded to bring about a revolutionary transformation in society originated in late 18th century France with the Jacobins, and is thus both quintessentially Western and modern, a fact lost on most VEOs plying the trade today.

Behaviorism, Mind, and the Zombie Problem

Emerging in the late 19th century, Behaviorism holds that meaningful states of mind are exhausted by states of behavior. The mind is the brain in its many different states, and behavior is a manifestation of brain states expressed via the body's central nervous system. For the positivist/behaviorist, most notably B.F. Skinner, there is nothing left to know about this mechanical human person. Theories of mind that posit the existence of a 'hard problem'–the difficulty that no physicalist explanation of mind does justice to subjective inner experience–are overcomplicating the mind-body dilemma. For this brand of Behaviorism there is no hard problem. The existential dimension of human experience is ephemeral and probably epiphenomenal. The range of experiences humans can undergo, if caused by brain states, probably do not cause anything else.

One among many of the chief criticisms of Behaviorism (disseminated in popular culture) is that it allows little way of distinguishing between a person and a Zombie. If a person's subjective internal mental states are exhausted by his or her behavior, the way is open for a skillful Zombie with no internal mental state to nonetheless qualify as a person. This dismissal of the effects of inner experience–for Behaviorism's critics the most quintessential aspects of being human–contributed to its usurping during the 'cognitive revolution' led by Noam Chomsky in the second half of the 20th century. For

Chomsky and others, however, the rejection of Skinnerian Behaviorism was not founded on its apparent immorality but because it was found to be among the worst examples of scientism.

The explosion of available empirical data rendering human behavior offered by the digital age has renewed debate about the controversies of Behaviorism and its role, not only within cognitive science but also as a new species of controversy within the legal academy, political, and social sciences. The cognitive sciences today incorporate a cross-disciplinary research agenda which includes the computational sciences, social psychology, philosophy of mind, neuroscience, and complexity and network theory. This renewed vigor combined with the eclecticism of the field means often we are left guessing at the epistemic assumptions on which these controversies are being discussed. When it comes to the application of AI in NSID affairs, epistemic agnosticism cannot be sufficient.

Behaviorism and Pseudo-scientific Laws

Seeking to legitimize political science, Behaviorism emphasized the study of human behavior as a quantifiable political-economic phenomenon and as such eschewed the traditionalist approach to politics as a moral and ethical and therefore normatively private matter. For the behaviorist, all behavior was political behavior–no special separation need be granted for the ethical realm. Political parties seeking to understand and influence the population found much value to be mined in the methods and means of Behaviorism. Voting behavior could be studied for its patterns and regularities much like economic behavior – if quasi-laws could be discovered, parties privy to this information could gain a decisive edge.

Propaganda as political advertising, campaign strategy, and targeted communications were all fruits of the rise of Behaviorism in politics, and as such shared a number of features with earlier trends in economics. The idea that the study of human affairs could be reduced to that of a systems science was introduced into economics by a group of scientists and philosophers known as the ‘Vienna Circle’ in 1907. Based partly on the work of Ernst Mach (1838-1916) and fusing with that of the Positivists, the Vienna Circle declared religion, metaphysics, and morality to be meaningless. Only propositions emerging from a singular vision of scientific method qualified as meaningful, and this declaration helped the ‘science’ of economics achieve vaunted elevation aside physics and mathematics. This marked the beginning of such things as the ‘laws’ of economics.

Comtean ideas, transmitted to the United States via the scattering of the Vienna Circle abroad after the war, had significant influence on neo-liberal economists including Milton Friedman and marked a definitive break from the work of classical economists such as Smith and Ferguson for whom history–not mathematics–was the basis of their understandings.

For American capitalism more broadly, the 1980s saw firms begin to see commercial competition differently from the past. Having primarily been a contest between firms to attract the consumer with superior products and services, commercial competition became a contest fought against the consumer for their patronage. Firms worked harder and devoted more capital to understanding the art of extracting money from the consumer’s hands than the contest between firms, which combined with other macro-economic shifts, had by the mid-1980s driven profit margins down across the board. Firms were driven by new ‘laws’ of commercial competition to find different vectors to success.

The art of consumer manipulation entered a new phase with the offshoots of neo-Behaviorism at its core. Parallel with the emerging digital age, the Internet, mobile connectivity, and big data analytics, capitalism underwent a mutation in the early 2000s, which is only recently becoming widely understood. Combining the historical trajectories of Positivism and Behaviorism with the saturation of the digital age and advances in AI, Shoshana Zuboff dubbed this new form 'Surveillance Capitalism' in 2015. Human behavior becomes the primary extraction resource necessary for the productive processes of surveillance capitalism. Its products are behavioral predictions, rendered by algorithmic learning technologies sold at auction to corporate bidders seeking a new level of access to, and certainty about, the consumer. For this rogue mutation, the extraction and denaturing of human behavior provides commercially lucrative insights into the new 'laws' of socio-political-economic interactions in the digital age.

It also raises fundamental new questions about the social contract, democratic legitimacy, and economic distribution, and leads us back to fundamental existential questions about human beings, mind, society, freedom, authority, knowledge, and power. These are questions not to be rushed into nor dismissed as philosophical luxuries. The NSID community needs to clearly demarcate its interests in AI technologies for the conventional battlefield, where they offer undoubted advances in discrete tasks, from other uses in society where they represent radically untested interventions in complex anthropological systems.

The popularity and profitability of behavioral economics does not mean its methods and assumptions transfer into the NSID space. In fact, the opposite may be more accurate. NSID affairs should approach these mutations in capitalism with caution because they represent a series of radical interventions into complex anthropological systems. Risk gives way to uncertainty, and the locus of contestation becomes the very social fabric upon which any notion of strategic gain is fundamentally dependent. 'Winning' in AI should not be pursued at such exorbitant cost. If our adversaries and competitors are willing to enter this zone, only hubris suggests we should follow.

Dualism and Some Ethical Sleight-of-Hand

Despite a number of noteworthy and continuing efforts globally, ethics in AI continues to lag behind practice. Detectable in this gap is a distinctive reversion to Dualism, a relapse in thought at least partly explainable as a result of the speed of change but also a sign of a fundamental human need. The need that Dualism is supplying is generated by the existential discomfort of the digital age. Science and technology (S&T) have, since the time of Socrates, generated existential discomfort for human beings–the age of AI is only the latest culmination point. The threat of S&T to the sanctuary of the human mind has, most famously since Descartes, generated all manner of metaphysical defenses.

Cartesian Dualism is a position within the philosophy of mind which holds that mind and body are either composed of fundamentally distinct *substances*; body physical and mind non-physical; or exhibit fundamentally distant *properties* which define their separation. Almost nobody within mainstream cognitive science today is a Dualist. Descartes famously tried (and failed) to explain how the two fundamentally distinct entities of body and mind could interact without being detectable–in other words–without sharing some ontological commonality. Despite its debunking, Dualism continues to exert influence. When the discussion in AI turns to the ethics of behavioral modification, its encroachment on liberty, and the erosion of the concept of free will, we see Dualism creeping back in as some in the AI community seek a trouble-free ethical justification for their research.

For the positivist/behaviorist, the detachment of inner subjective human experience from outward behavior is itself a form of Dualism. Whether explained away as epiphenomenal, or taken as ephemeral and therefore inconsequential, the question of subjective inner human experience is left unaccounted for by Behaviorism. From here it is an easy step to cast the questions of liberty and free will as unimportant in the science of behavioral modification. The behaviorist need only assert they are working on the side of 'freedom and well-being,' or any other socially and politically acceptable platitude, and there is nothing left to justify. Unfortunately, a number of significant problems underlie this sleight-of-hand.

It is not sufficient to simply declare the 'hard problem' inconsequential and to eschew further debate, while taking forward the methodologies and assumptions of Positivism and Behaviorism. The assumptions of Positivism and Behaviorism are far from inconsequential–they are not value-neutral. They accommodate a set of orientations toward the nature and origins of reality, and the existential place of human life within that schema, which represent currents in the history of European thought and trends in contemporary science. These currents and trends will continue to change, sometimes radically. Notwithstanding pockets of consensus here and there, their controversies are not settled, nor should we expect them to be. When interventions in complex anthropological systems are considered on the basis of these assumptions, our approach should be cautious and open to the possibility our own misconception–not cavalier. This is the strategic essence.

Conclusion

Without always understanding it, and often without stating it, the majority of research and development in AI S&T inherits both its epistemology and its methodology from Positivism. Saint-Simon and Comte envisioned the convergence of all human knowledge into a single set of laws, applicable across the spectrum of human affairs. The 20th century rebuffed their hopes–the digital age has renewed them for a new generation. In its narrow focus on the extraction via digital rendering of human behavior and the search for hidden patterns, Behaviorism underpins the approach to the mind-body question often left unstated in AI S&T. Behaviorism dominates the lucrative and highly controversial commercial and political practices we associate with Surveillance Capitalism, and Dualism has slipped back in as a commonly deployed moral and ethical equivocation.

Militaries rightly focus on the discrete and bounded problems and tasks AI can expedite, but they are also tempted, with the rise of cognitive and population-centric warfare, to utilize some of AI's potential to seek advantage in these indiscrete and unbounded spaces as well. This temptation must be resisted.

We should consider some fundamental questions as the 'AI race' continues. What becomes of our concept of the person, the good society, and the nation as a socio-cognitive construct in this stampede? What does 'winning' mean? The extraction of behavioral surplus for commercial gain, and for the advancement in AI learning it has fed over the past twenty years, has had a demonstrable impact on the social fabric in every country where it has been practiced. Contrary to the tenets of the thought trajectories examined here, any sustainably adaptive social fabric is composed of threads that extend beyond our capacity to observe, quantify, and responsibly manipulate. Human systems are open, nonlinear, and complex, and while complexity theory contains sub-theories offering ways to cheat chaos by altering initial conditions, relying on path-dependencies, and avoiding system perturbations and so forth, a great deal of caution should be wielded.

Intervening in human complexity involves existential uncertainty, not just instrumental risk, to which technology adds a powerfully protean element. When considering these possibilities, it pays to remember the assumptions on which they are predicated are little more than iterations in the history of European thought, rather than arrows to truth. As John Gray writes, "For Saint-Simon and Comte, technology meant railways and canals. For Lenin, it meant electricity. For neo-liberals, it means the Internet. The message is the same. Technology–the practical application of scientific knowledge–produces a convergence in values. This is the central modern myth, which the Positivists propagated and everyone today accepts as fact" (Gray, 2007).

The danger of such a widely-accepted fact is that we cannot deter, let alone defeat, the threats we cannot imagine. Hubristic interventions in complex anthropological systems, which disrupt the basis of their functionality in ways not captured by our current best models, might be one such threat.

Annotated Bibliography

Ed. Note: The following bibliography includes the sources cited in the text, as well as other related works that are relevant to the points made in the paper. Annotations are designed to serve as a guide for further study, and this bibliography is not intended to be an exhaustive list of related works.

Bi-modal/"Barbell" strategy

Taleb, N. N., Bar-Yam, Y., Douady, R., Norman, J., & Read, R. (2014). The Precautionary Principle: Fragility and Black Swans from Policy Actions. *Extreme Risk Initiative - NYU School of Engineering Working Paper Series*, 1–20. Retrieved from http://www.fooledbyrandomness.com/precautionary.pdf.

Taleb, N. N. (2014). *Antifragile: Things That Gain from Disorder*. New York: Random House Trade Paperbacks.

Historical readings of philosophy

Bourdeau, M. (2008). Auguste Comte. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018). Retrieved from https://plato.stanford.edu/archives/sum2018/entries/comte/

Manuel, F. E. (1956). *The New World of Henri Saint-Simon*. Harvard University Press.

Manuel, F. E. (1962). *The Prophets of Paris: Turgot, Condorcet, Saint-Simon, Fourier and Comte*. Harper & Row.

Mill, J. S. (1968). *Auguste Comte and Positivism*. Library of Alexandria.

Pickering, M. (1993a). *Auguste Comte: Volume 1: An Intellectual Biography*. Cambridge University Press.

Pickering, M. (1993b). *Auguste Comte: Volume 2: An Intellectual Biography*. Cambridge University Press.

Stadler, F. (2015). *The Vienna Circle: Studies in the Origins, Development, and Influence of Logical Empiricism*. Springer International Publishing.

Philosophy and Technology

Gray, J. (1995). *Enlightenment's Wake: Politics and Culture at the Close of the Modern Age*. London: Routledge.

Hoffman, D. D. (2008). Conscious Realism and the Mind-Body Problem. *Mind & Matter*, *6*(1), 87–121. *Ed. Note: Hoffman provides a summary of contemporary attempts at the 'hard problem'*

Osinga, F. P. B. (2007). *Science, Strategy and War: The Strategic Theory of John Boyd*. Routledge.

Zuboff, S. (2015). *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization* (SSRN Scholarly Paper No. ID 2594754). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2594754

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile.

'Nudging': Arguments for and Against

Afif, Z. (2017, October 25). "Nudge units" – where they came from and what they can do. Retrieved September 11, 2019, from Let's Talk Development website: http://blogs.worldbank.org/developmenttalk/nudge-units-where-they-came-and-what-they-can-do

Sunstein, C. R. (2009). *Going to Extremes: How Like Minds Unite and Divide*. Oxford, New York: Oxford University Press.

Thaler, R. H., & Sunstein, C. R. (2012). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin UK.

Wright, J., & Ginsburg, D. (2012, February 16). Free to Err?: Behavioral Law and Economics and its Implications for Liberty. Retrieved from Law & Liberty website: https://www.lawliberty.org/liberty-forum/free-to-err-behavioral-law-and-economics-and-its-implications-for-liberty/

Ed. Note: Sunstein, as well as Wright & Ginsburg's work offers counterarguments against the principle, whereas the joint work by Sunstein & Thaler, along with Afif present arguments for.

Positivism: classical and modern approaches

Comte, A. (1856). *Social Physics: From the Positive Philosophy of Auguste Comte*. C. Blanchard.

Pentland, A. (2010). *Honest Signals: How They Shape Our World*. MIT Press.

Pentland, A. (2014). *Social Physics: How Good Ideas Spread — The Lessons from a New Science*. Scribe Publications Pty Limited.

Pentland, A. (2015). *Social Physics: How Social Networks Can Make Us Smarter*. Penguin.

Ed. Note: Pentland's included works represent a contemporary application of Positivist thinking.

B.F. Skinner

Chomsky, N. (1971, December 30). *The Case Against B.F. Skinner*. Retrieved from https://www.nybooks.com/articles/1971/12/30/the-case-against-bf-skinner/

Skinner, B. F. (1966). *Walden Two*. Hackett Publishing.

Skinner, B. F. (1972). *Beyond Freedom and Dignity*. Bantam Books.

Ed. Note: Chomsky's work is included as a critique of Skinnerian Behaviorism.

Ethics and Artificial Intelligence

Floridi, L. (2013). *The Ethics of Information*. OUP Oxford.

Floridi, L. (2018). Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philosophy & Technology*, *31*(3), 317–321. https://doi.org/10.1007/s13347-018-0325-3

Giordano, J., & Gordijn, B. (Eds.). (2010). *Scientific and Philosophical Perspectives in Neuroethics*. Cambridge University Press.

Definitional terminology

The World Bank. (2015). *World Development Report 2015: Mind, Behavior, and Development*. Retrieved from: http://documents.worldbank.org/curated/en/645741468339541646/pdf/928630WDR0978100Box385358B00PUBLIC0.pdf

National Security Implications

Gray, J. (2007). *Al Qaeda and What It Means to Be Modern*. London: Faber & Faber.

Lawson, S. T. (2013). *Nonlinear Science and Warfare: Chaos, Complexity and the U.S. Military in the Information Age*. Routledge.

United States Department of Defense. (2019, February 12). SUMMARY OF THE 2018 DEPARTMENT OF DEFENSE ARTIFICIAL INTELLIGENCE STRATEGY: Harnessing AI to Advance Our Security and Prosperity. Retrieved September 11, 2019, from https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF