Before Eschaton Lies the Wor(l)d

U. Kanad Chakrabarti

2019

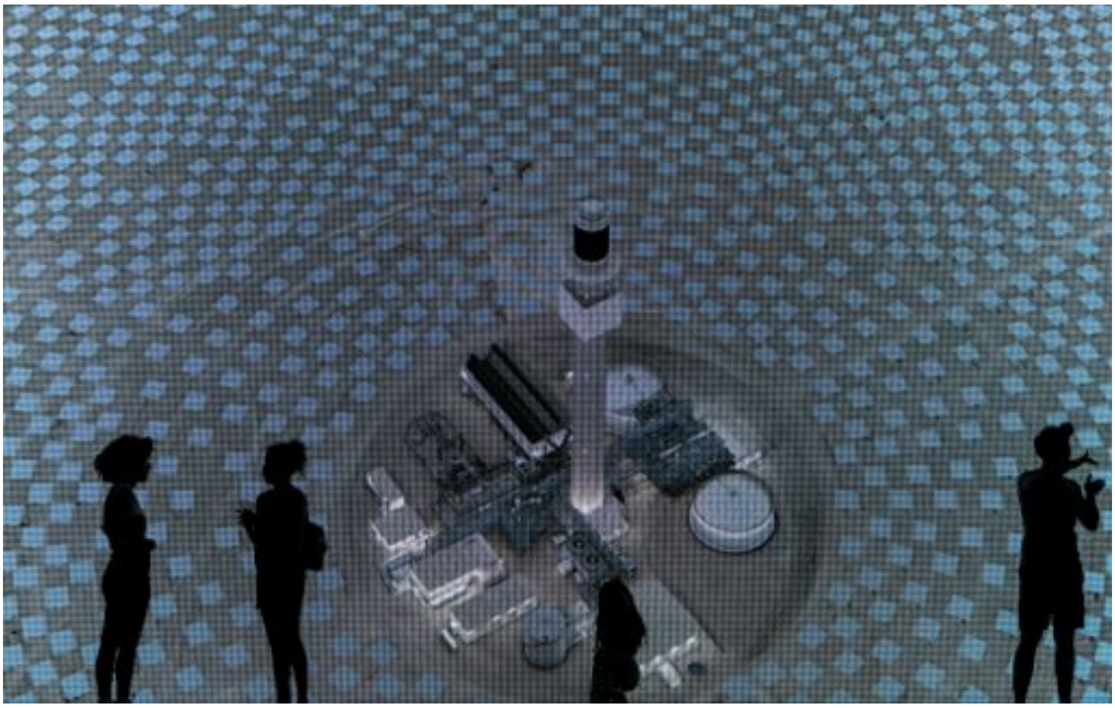

Figure 1: John Gerrard *Solar Reserve (Tonopah, Nevada)*, 2014, Simulation. Source: johngerrard.net

> *'The stone is world-less. Similarly, plants and animals have no world; they belong, rather, to the hidden throng of an environment into which they have been put. The peasant woman, by contrast, possesses a world, since she stays in the openness of beings.'*

Heidegger, Martin. *The Origin of the Work of Art*, 1960, p.23

To create new worlds or speculative imaginaries has, for some years, been a fertile tract of contemporary art, for instance in the work of Ian Cheng, John Gerrard and Pierre Huyghe. Indeed, these very different artists offer mirrors, sometimes faithful, and others distorting, to a contemporary society that is characterised by ever-shifting identities, a continuous and spectacular exchange, and modes of being that increasingly dissolve the man-machine limen.

'Worlding', coined in Martin Heidegger's distinctive, gerundive diction, takes on new life in the algorithmic contemporary, by allowing us to create environments within the machine, and increasingly, within networked environments, such as an exhibition space or online. This essay unpicks the thread of simulation, with its proto-history in the thought-experiment, that meets modernity during the 1950s development of thermonuclear weapons, and now plays a central role in Artificial Intelligence (AI) research.

Heidegger, Martin. *Being and Time*, 1927

Computer simulation obviously has wider use than in nuclear weapons and AI. For instance, the derivatives pricing revolution of 1990s, as well as weather forecasting, form fascinating stories, yet must be left for a future essay.

Specifically, through simulation we have gained the ability to quantify contingency: by allowing for thousands or millions of scenarios to be generated, at varying temporal or spatial scales, computer simulation allows both man and machine to 'play out' alternative futures. For instance, Google DeepMind's recent successes with the Japanese board game Go has been a much-touted victory for AI. Go, unlike chess, cannot be mastered at a world-champion level by brute-force methods (such as calculating all possible permutations of moves). The search space, on the order of 10^{170}, is simply too large, and it is necessary for a player, human or cybernetic, to wisely choose subspaces that, with high probability, contain winning strategies.

DeepMind's algorithms, described in detail here and visually summarised here, are trained over thousands of possible games and utilise a method called Monte Carlo Tree Search (MCTS), balancing a broad search through the space with in-depth evaluation of promising move sequences, a strategy known as 'exploration versus exploitation'. The algorithm conceptually is a cyclical process of intelligent guesses at future moves, simulation of those moves, consequent assessment of the value of that move/guess, followed by an updated assessment of the game's state. It is notable that the algorithm, training itself by playing itself, invented new tactics of great power, that human master-players found surprising and counter-intuitive.

MCTS is based upon Monte Carlo simulation, a numerical technique developed in the 1950s by researchers tackling the challenge of the thermonuclear or fusion bomb. To see why simulation was necessary, it is worth digressing briefly to examine the differences between a fusion and fission bombs. A crude fission or atomic bomb can be constructed, and tested, by explosively compressing a sphere-shaped critical mass of plutonium. While the engineering challenges of machining the necessary pieces were significant, as were the logistics of procuring and refining fissile material, the basic concept and design were not a mystery. In contrast, the thermonuclear bomb's design was considerably more complex, not least because (in the 'staged' design ultimately adopted by the US and USSR), a small fission weapon (the 'primary') would be used to ignite a larger 'secondary' charge of hydrogen-based fusion fuel. In turn, this thermonuclear burn would, in some variants, set off a third fission stage. The multiple reactions within this staged design would need to happen within a few nanoseconds, the time-scale over which the weapon itself dis-assembles, or breaks apart, under the tremendous heat and pressure of the primary fission detonation. As historian-of-science Peter Galison writes:

Galison, Peter. *Computer Simulations and the Trading Zone* in *From Science to Computational Science*, edited by Gabriele Gramelsberger, 118-157. Zúrich: Diaphanes, 2011, p. 122

> *'Not only were the nuclear physics of hydrides, the diffusion of hard and soft radiation, and the hydrodynamics of explosion difficult in themselves; they had to be analysed simultaneously and in shock at temperatures approaching that of a stellar core?.Experiments seemed impossible - a hundred million degrees Kelvin put the laboratory out of the picture; there was no thermonuclear equivalent to Fermi?s reactor, no slow approach to criticality obtained by assembling bricks of active material.'*

Ian Goodfellow and Yoshua Bengio and Aaron Courville. *Deep Learning*, MIT Press, 2016, p. 152, http://www.deeplearningbook.org

In addition to the obvious inaccessibility to experimentation, thermonuclear weapons designers needed to model the movement of a myriad of particles in a myriad of potential positions, resulting in an enormous space of possibilities (akin to what, in an AI context, is termed the 'curse of dimensionality') that cannot be navigated via standard closed-form equations, but are not amenable to traditional statistical mechanics either. Rather a new approach had to be developed that essentially simulated the life of each particle, tracking its path, its collisions, and whether it contributed to the desired chain reaction. The result of each simulated particle-lifetime would be tallied, and the whole lot summed up using an appropriate weighting-function, giving the final result - an estimate of whether the weapon was likely to reach a self-sustaining reaction, i.e. a detonation. This calculationally intensive process, while repetitive, can be algorithmically specified, and was well-suited to the room-sized computing machines that were coming on-stream in the 1950s.

Jacques Derrida, Catherine Porter and Philip Lewis, *No Apocalypse, Not Now (Full Speed Ahead, Seven Missiles, Seven Missives)* in *Diacritics* Vol. 14, No. 2, Nuclear Criticism (Summer, 1984), pp. 20-31, available here, last accessed January 2019

Alongside the the numerical simulations above, Cold War nuclear planning also gave rise to a practice of qualitative scenario analysis. Nuclear apocalypse itself had (and still has) a perversely hypothetical, literary quality, held in the collective imaginary, what Jacques Derrida termed the 'fabulously textual'. Think-thanks such as the Hudson Institute or the RAND Corporation used scenario analysis to explore the ways a nuclear conflict could play out. They posed questions like: could an adversary survive a first-strike long enough to launch a counter-strike; could a nuclear war be won; how might the nation re-build its economy after an less-than-entirely successful attack? These questions underpinned, and arguably undermined, the game-theoretic doctrine of mutually-assured destruction (MAD) and the perceived stability of the US-USSR strategic relationship.

Figure 2: Herman Kahn (L) advising President Gerald Ford and Donald 'Unknown Unknowns' Rumsfeld on how best to win World War III. Source: dinmerican.wordpress.com

Herman Kahn, founder of the Hudson Institute and an inspiration for Stanley Kubrick's Dr. Strangelove, saw the apocalyptic scenario as essentially a story, intensely specific, the parameters of which were asymptotic - or using Galison's terminology again, a caricature. This caricature, by virtue of exaggeration and clarity, would force war planners to assess, analyse, and respond to the complex interaction of military, economic, social, and human factors that might come into play within that fateful half-hour interval between a Soviet launch and annihilation.

Kahn's scenarios, like simulation, were conceptually indebted to the philosophical thought-experiment, a tool with a rich history, from Plato's celebrated cave onwards to Schrödinger's cat to Searle's Chinese Room. A series of 'what-if' statements, couched within a precisely-defined problem domain, and based upon parsimonious axioms, allow for speculation about questions that cannot, for reasons of complexity, scale or sheer destructiveness, be subjected to experiment. Computer simulation could, given its aims, also be seen as thought-experiment *in silico*. Yet there are important differences: for instance, simulation generally tries to emulate physical, economic, or biological processes to find a precise, numerical answer (as in the nuclear case above), which can be subjected to traditional tools of statistical analysis (as in a standard physical experiment). In contrast, the thought-experiment is often more qualitative in nature, and to the extent it is quantitative, it tries to seek numerical bounds or limits. The thought-experiment also tries to the think around the problem, re-defining it using ingenious logical stratagems (for instance, Cantor's 'diagonal slash' that treats the sizes of infinite sets). The simulation, in contrast, is generally confined to a specific problem. Importantly, the thought-experiment, by virtue of its categorical breadth and syntactic flexibility, allows for the analysis of ethical, logical or aesthetic considerations.

Penrose, Roger, *The Road to Reality*, (2004), pp. 364-365.

The thought-experiment and computer simulation come together in Oxford philosopher Nick Bostrom's research. In a 2003 paper Bostrom analyses the intriguing possibility that we (humans alive now) are living within a simulation created by our own sufficiently-advanced, or

Nick Bostrom, 'Are You Living in a Computer Simulation' in *Philosophical Quarterly* Vol. 53, No. 211, pp. 243-255.

'post-human', descendants. This simulation, Bostrom hypothesises, is running on a 'computer' that is, architecturally and algorithmically, far beyond our own capabilities or imaginations. Bostrom casts this thought-experiment into a fairly simple formula that, assuming the logical and numerical assumptions underlying it are not found wanting, partitions reality (that which underlies our subjective experience) into three possibilities, one of which is that we are living within a simulation. The alternatives, both of which would force us to conclude that we are not living within a simulation, are: (a) any such (hypothetical) post-human civilisations are likely to have destroyed themselves in an extinction event (such as a nuclear war or a malevolent AI) before they have had the chance to conduct a simulation of their ancestors; or (b) such post-human civilisations, even if they are capable of it, for one reason or another, have no desire to simulate their ancestors. Bostrom, in considering possibility (a), has carefully enumerated over thirty varieties of existential apocalypse. Drawing upon the thought-experiments of another Oxford philosopher, the late Derek Parfit, Bostrom tries to quantify the 'loss' associated with human extinction; callous as it might seem, we need, from a public policy perspective, such a number to assess what level of current (and scarce) resource allocation is appropriate to guard against this eventuality. As far as alternative (b), hypothetical post-humans might, upon reflection, conclude that simulating their ancestors is fraught with ethical issues - for instance, the suffering (death, illness, inequality) that would be inflicted on said subjects in the simulation.

Roussel, Raymond *Locus Solus* (1914).

Bostrom's ideas are admittedly speculative, but clearly not irrelevant. Although the technologies he posits appear to be a distant possibility, what do his conclusions mean for such semi-autonomous intelligences, 'ghosts within the shell' pre-figured by Raymond Roussel, and for instance, alluded to in the work of Ian Cheng, that we might eventually create? Namely, what, if any, ethical duty of care might we have to these creations? Moreover, pace Bostrom, once we can 'upload' our minds into some sufficiently advanced computing system, does this potentially make us immortal, infinitely replicable, and resilient to physical damage? What does this mean for the angst of existence that comes from awareness of looming death, from entropic decay in mind and body, that is so embedded in our genetics, economics, social organisation, and culture?

There are some more concrete, immediate implications of Bostrom's thought. For instance, in assessing the likelihood of ancestor simulation, Bostrom considers the 'cost', in energy or resource terms of building a suitable computer, which could be tremendous or even prohibitive. Metaphorically, this highlights a major concern in AI - algorithms are trained upon vast amounts of human-generated data, which are (in-)voluntarily donated by users of social-media, mobile phones and online services. In Western democracies, this comes at the cost of privacy, entrenched corporate power, and disruption of political norms. In China, the threat to human agency is arguably worse.

The point of diving, at some length, into Bostrom's ideas is to highlight the ethical possibilities that separate thought-experiments from computer simulations. The former, being an essentially humanistic activity, allows us to 'take back control', as it were, from an overly techno-scientific focus that myopically, or even intentionally, avoids normative judgements. This is particularly important when said un-checked technological progress is, as it were, yoked to socio-economic structures - neoliberal capitalism or totalitarian regimes - that are inherently inimical to emancipatory ideals.

Heidegger, Martin *The Question Concerning Technology* (1977).

> *'Everything depends on our manipulating technology in the proper manner as a means. We will, as we say, "get" technology "spiritually in hand." We will master it. The will to mastery becomes all the more urgent the more technology threatens to slip from human control.'*

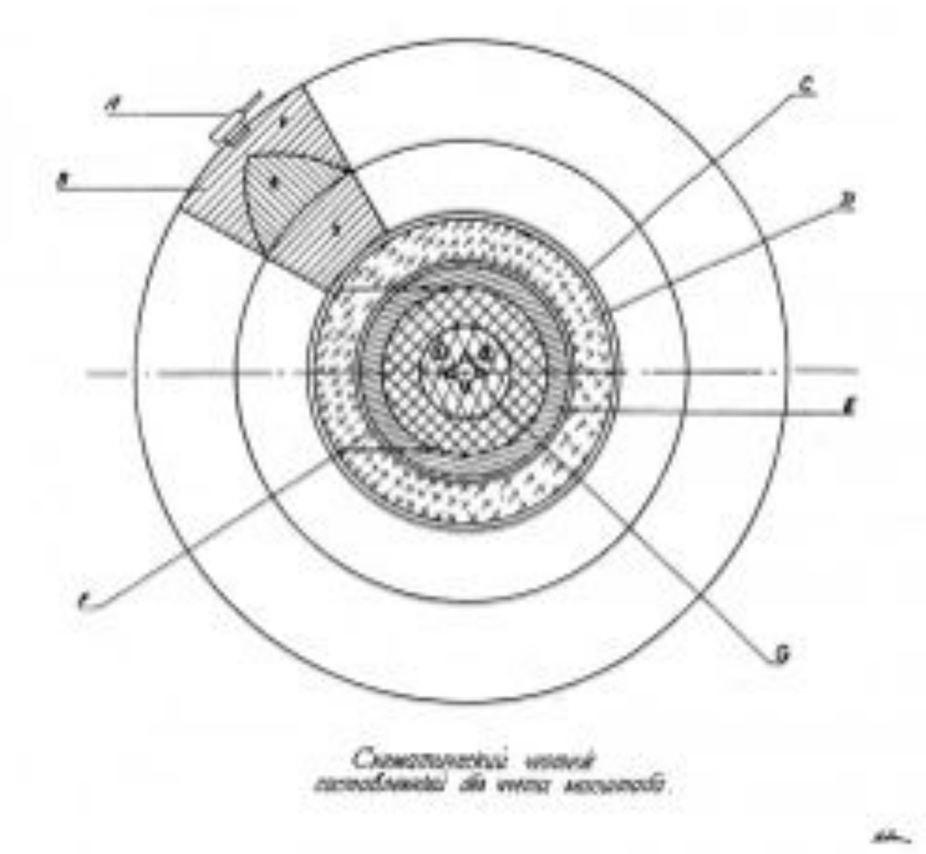

Figure 3: Soviet schematic of a Trinity-style implosion weapon showing explosive lenses. The behaviour of plutonium, and other materials, under compressive shock is notoriously difficult to model without high-performance computers and simulation. Source: armscontrolwonk.com