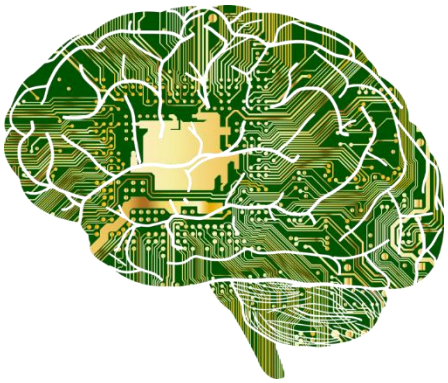
Blog of the APA

A Neurophilosophy of Two Technological Game-Changers: Synthetic Biology & Superintelligence

By Nayef Al-Rodhan

June 29, 2020



This is post six in a short-term series by Prof. Nayef Al-Rodhan titled "Neurophilosophy of Governance, Power and Transformative Innovations." This series provides neurophilosophical perspectives and multi-disciplinary analyses on topics related to power and political institutions, as well as on a series of contemporary transformative technologies and their disruptive nature. The goal is to inspire innovative intellectual reflections and to advance novel policy considerations.

While many of the experiments in the domain of human genetic modification and synthetic biology remain at an early stage, there can be no doubt that they put humanity in uncharted territory. Barriers to much more dramatic forms of human enhancement—by way of genome editing and synthetic biology, and in the field of *superintelligence*—continue to fall away as the relevant scientific fields advance. Amidst this furious pace of development careful consideration must be given to the place of humanity in these developments, along with the social, political, ethical and existential implications of these pursuits.

A neurophilosophy of these dramatic technological developments can help to contextualize human agency and responsibility in light of them. Acknowledging our own neurochemical makeup and the social and political tendencies it motivates offers insight into the kind of policies that should be taken in overseeing the development and dissemination of technological gamechangers. This is particularly true with respect to super intelligence, given its potential to evolve and metastasize in directions outside human control.

Technological Innovation and Human Nature

Synthetic biology

Synthetic biology is a very recent interdisciplinary field – from around 2000– although its origins go back to the late 1970s. The definition of synthetic biology remains somewhat contested as some critics consider it "extreme genetic engineering." However, at the core of synthetic biology is not merely to

alter biological systems, but *to design and create novel biological systems with functions that do not exist in nature*.

Succinctly, there are three main lines of work in synthetic biology:

1. Synthetic genomics, which refers to the development of simplified microorganisms that can be later used for further alterations (for example, in 2010, J. Craig Venter Institute announced the first synthetic organism – resulted from inserting a synthesized version of *Mycoplasma mycoides* into a cell body of *Mycoplasma capricolum* which subsequently turned into a fully functional cell with the attributes of *M. mycoides.*)

2. "Biobricks" or "bioparts" defines the efforts to create simplified and standardized genetic sequences that cause a microorganism to perform very specific tasks

3. "Novel biochemistries" – a group of research endeavors aiming to create "protocells" (which may have their own mechanisms and metabolism); this group of work also includes research in what is known as "mirror life", a line of research that works to produce entirely novel genetic creations based on mirror-images of genetic material.

The ultimate goal of applying engineering principles to biological systems is to serve human ends, and this includes applications for industry and society, such as new processes in agriculture, new medicines, environmental remedies or the production of new biofuels. For instance, some synthetic biologists imagine something similar to an 'off' switch for engineered plants so they could respond to environmental signals, such as dryness. If carried out safely and responsibly, such innovations could help tackle human and global security challenges (e.g. food crises, fighting pollution). However, synthetic biology is raising a host of environmental concerns as it implies releasing *synthetic* organisms into the environment, which makes it crucial to design not only robust but also highly *predictable synthetic organisms*.

The more extreme and revolutionary innovation in synthetic biology is, however, increasingly occurring in relation to the human body. The principles of synthetic biology are now studied, among others, in the human microbiome, exemplified by the work on a synthetic probiotic, in the form of a "bacterial species that detects and treats diseases from inside the gut". Successfully engineering such a bacterial system could, in theory, help the body in a host of healing processes (and now we know the gut microbiome plays a fundamental role in health, including in the manifestation of some mood disorders, anxiety and depression). This synthetic probiotic system would colonize the gut, diagnose a disease state and initiate 'outputs,' meaning production of therapeutic genes, or alteration of signaling molecules etc.

Other important ventures of synthetic biology are already testing the line between treatment and human enhancement – hailing an era of very consequential developments both in society and for the military. An emerging field of application is within transplantation, in at least two ways: through tissue engineering in vitro and in vivo to create organs, and through the use of genome engineering to address the critical issue of host rejection. Furthermore, it is hoped that the rapid developments in synthetic biology could help design organs that "eventually surpass human organs in function and survival."

Advances in synthetic biology and the human genome have progressed on even more daring fronts, including in the field of synthesizing human DNA, carried out by scientists in the Genome Project-Write project (or GP-Write). Tools for genome editing that are premised on cutting and pasting fixes in the human DNA have become well-known following the groundbreaking innovations of technologies such as Crispr. GP-Write, however, aims not only to 'edit' DNA but to re-write critical stretches of chromosomes, which can subsequently be stitched together with a naturally occurring genome. One scientist in the project explains the difference between gene editing with Crispr and synthesizing modified versions of genes as "the difference between editing a book and writing one". In more concrete terms, this means creating synthetic replicas of an individual's sequence but recoding it, for example, to be resistant to viruses. The process of "recoding" means that select nucleotides in the chromosome will have their sequence changed. (The scientists behind the project insist, however, that creating new types of humans or babies is *not* their ultimate goal, 'just writing genomes'). In 2017, the collaborators in GP-Write held a public meeting of over 250 participants, including lawyers, policymakers, and ethicists – signaling the wide interest and concerns over the ambit of projects that aim to manipulate the code of life.

Achieving anything close to that is however, still a relatively distant possibility seeing how complex a process such as synthesizing even parts of the human genome is in reality. Nevertheless, even if this process may be very slow, the promises of new technologies in life sciences signal important changes for humanity, both for our biological future and our cognitive abilities. The idea of superintelligence, which has thus far been described as hypothetical machine intelligence that surpasses the smartest human mind, may acquire new meaning as possibilities in synthetic biology advance – and with these, the feasibility of biological superintelligence.

Superintelligence

Superintelligence – the stage of an intelligent entity that surpasses the greatest possible cognitive abilities of a human being – has generally been theorized in relation to technology.

In *future studies* scenarios, superintelligence is believed to follow shortly after the advent of general artificial intelligence, as systems of logic and reasoning attain a form of intelligence no longer bound by the limitations of human cognition. This includes, among others, perfect memory, unaltered by age or any circumstances, the ability to develop skills and competencies in a vastly wider array of fields, the capacity to multitask (still very difficult for the brain), or significantly heightened vision or other sensory abilities. Once achieved, superintelligence could unlock mysteries and knowledge about the universe, afford the possibility of eternal life or human intergalactic journeys, and open limitless breakthroughs in all fields of science. (The very existence of superintelligence may, however, threaten human existence to the point that it will aim to dominate or even exterminate us. I will not dwell on this aspect here, however.)

Quite ironically, one source of inspiration for superintelligence is the human brain itself. Modern neuroscience is still far from having decoded the entire brain but its workings are known in at least some fundamental ways; among the things we know is that it relies on the existence of complex neural networks, learning happens through reinforcement, and that it has a hierarchical structure. Some of the imagined paths to superintelligence have drawn on such insights into the human brain and neural networks but synthetic biology is poised to lead to new approaches that shift the focus to *biological superintelligence*. Therefore, in addition to relying on artificial intelligence, super-computers or 'whole

brain emulation' (a process also known as 'mind uploading'), and synthetic brains, the goal of super-intelligence may be considered and fostered within the human mind and through radical changes to human biology.

Speculative accounts of the evolution of human intelligence in light of new bio-technological advances and genetic interventions are at least half a century old. Some hypotheses proposed in this debate included views on selective reproduction and ways to permit our species to grow larger heads, and will have surely seemed taken from sci-fi scripts back in the 1907s. Synthetic biology is still far from reaching its full potential, but it now offers more realistic glimpses of the range of transformations that can allow us to modify biological functions and – one day – design significant new capabilities into our cognitive capacities. Especially as knowledge about intelligence advances, including neuroscientific research on differences in human intelligence and research into the genetic understanding of intelligence, the field of synthetic biology will have a wealth of material to work with and build upon.

What does this mean for human nature?

The applications of synthetic biology to enable superintelligence are sure to cause profound shocks to humanity. Some of the risks associated with such extraordinary enhancement of biological and cognitive functions have been framed within the human enhancement debate (discussed in other posts here and here), but there are additional existential threats to consider.

There are important implications to the theory of human nature I have developed—*Emotional, Amoral, Egoism*—when applied to technological advancement. It is first important to note that genome editing and interventions of synthetic biology are both framed by and, to some unknown extent, have the potential to alter this account of human nature.

Moral concepts are not encoded in our genetic endowment, but we are innately programmed with some pro-social affinities, which provide a fertile ground for inculcating pro-social behavior and system of moral rights and duties. The emergence and consolidation of pro-social behavior, and social cooperation, are closely dependent on good governance and on human dignity – in the absence of the latter, the 'worst,' least cooperative and, at times, pre-emptively aggressive facets of human nature emerge.

The implications warrant further exploration. The first implication is that the process of developing these technologies is vulnerable to all the vicissitudes of human nature. This includes the possibility they might be deliberately harnessed by the powerful to subvert the weak, or to play out raw survival instincts in a variety of ways.

Alternatively, they might be developed in the spirit of enhancing global cooperation, limiting global threats, and pursuing goals like a more egalitarian world. However, which of these tendencies ends up being the case will largely depend upon the favorable or unfavorable social, cultural, and educational experiences of developers.

To put it differently, it is reasonable to suppose that super-intelligent entities, designed at least initially through human inputs, will be neuro-psychologically complex and will not be normatively neutral. An important realization for policy-makers is that they must focus not merely on final outputs—what firms are allowed to develop, for example—but also on the provision of quality universal education and social solidarity, which will inform final outputs.

The second implication is that the mere existence of a modified super-intelligent being would have re-authored human nature and would have the autonomous capacity to do so further. Grappling with the possibilities of fully autonomous AI developed to surpass human cognition many times over requires a strong sense of humility.

It is difficult to know how to program entities for which we have no precedents. A super-intelligent being, even if programmed via human understanding to have very low levels of anxiety, to be cooperative, stability-seeking, and so on, might nonetheless as a consequence of its greater epistemological scope and processing speed perceive itself to be in a Hobbesian state of nature, and thus act in exactly the hostile and preemptive ways its programming was intended to rule out.

Moreover, iterated creations of fully autonomous AI, by fully autonomous AI, clearly exacerbates these issues of unpredictability. Because we can at best make educated guesses with respect to how super-intelligent entities would behave, there is all the more reason to restrain its creation, simply because the implications are not well understood to assess their associated risks.

Politics, Neurochemistry, and Superintelligence

Given these considerations, there are profound political impacts to the development of synthetic biology and super-intelligence. All the familiar concerns with power, and the instinct to accrue, preserve and maximize power recur in new and troubling ways.

As I have argued elsewhere, when left unchecked—whether by positive socialization or institutional constraints—the egoistic character of human nature will entail a relentless search for the consolidation of power, and for expanding it to the highest degree possible.

This is applicable to all walks of life including, political, executive, business, sports, academic, and other forms of power. These tendencies are not mere conjecture and are far from coincidental; the pleasure centers of the brain have dopamine receptors that respond positively to various stimuli and encourage the repetition of whatever behaviors that produce this chemical.

Feelings of dominance and emotions associated with overcoming competition, are among the known experiences that tend to produce dopamine, serotonin, endorphins, oxytocin, or similar positive-reinforcement chemical release. In "normal" human contexts, these facts lead to healthy competition but can readily lapse into the phenomenon of power addiction, and often an associated collapse of legitimate governance into unaccountable and divisive rule.

The prospect of super-intelligence dramatically heightens these risks. There is every reason to think that the increased power conferred by being biologically enhanced and super-intelligent would amplify issues of power addiction, so that the augmented elite or even the most augmented individual would prioritize their preservation and increase its own power above all else.

A super-intelligent being, furthermore, would have greater opportunities to more successfully manufacture consent, remove checks on its power, and engage in various activities the political theorist Philip Petit describes as "domination." Indeed, the problematic of domination is a useful framework through which to consider a number of the conceptual issues posed by super-intelligence. A central point in Pettit's arguments is that the achievement of real freedom presupposes a certain kind of unassailable equality.

This argument has ancient roots, of course, and goes to the essence of questions regarding human nature and developments in human augmentation. In his elaborate taxonomies, Aristotle infamously inferred something akin to an "ought" from an "is" when he concluded that humans fundamentally were as they were coincidentally found. No changes of class were possible – a narrative that fundamentally canonizes the unbridgeable inequality between certain classes. Rousseau provided his concise response in 1762, describing the attitude of Aristotle as half correct, and what we might describe as negative socialization.

Rousseau articulated this understanding in a way that—like the arguments proffered by Mary Wollstonecraft with respect to the condition of women—signals a particularly modern understanding of human plasticity and the critical impact of socialization. For Rousseau, this realization was intimately related to another: the kind of "equality" his social contract theory forebearers claimed to exist naturally was in fact established by convention, if it was to be had at all. Equality was best approximated by equal treatment, and equal treatment was possible only via conventions and institutions that demanded it be afforded to all citizens.

Designing our common future: the way forward

The achievement of the 'Singularity,' the creation of a true super-intelligence, would throw the contours of this entire intellectual history into disarray. While there is much to admire in Rousseau's understanding, particularly as a prelude to the Rights of Man, his central arguments rest upon his indictment of Aristotle, and the denial that humans come in different grades, even if political equality can only be achieved by convention.

Particularly in the case of a super-intelligence achieved through synthetic biology, Aristotle's arguments would be precisely accurate. The superiority of such entities would be traceable right to their molecular makeup, as would the inferiority of the unenhanced masses—in both cases this being literally written into their DNA. For Rousseau, the great bargain of the social contract is to give up those freedoms fit for animals and the untutored, and to gain all the political freedoms associated with citizenship.

The frightening question that must be asked with respect to superintelligence, and one which should give pause in allowing its development to go further, is what the social contract has to offer an entity who not only judges itself to be but *actually is vastly superior* to its other members.

While the underlying premise of synthetic biology, as it is presented today, is to 'serve human ends', any 'ethically-unchecked' transgression into life sciences and human cognition carries profound risks for humanity and for all aspects of social cooperation. As always, it is a 'dignity check' that can make the difference between benefits and existential risks. Dignity, as I have theorized before, means much more than the absence of humiliation, and it includes a comprehensive set of nine needs: *reason, security, human rights, accountability, transparency, justice, opportunity, innovation,* and *inclusiveness.*

Each and every one of these needs must be preserved to prevent advances in cutting-edge technologies from creating alienation and irreparable damage to individual and collective dignity. One way to preempt this would be by bring scientists, ethicists, civil society organizations and policymakers together to agree on ethical and safety standards. If we are to move ahead with such powerful

technologies, this should at the very least be the result of wider deliberation in society. That too, however, could be insufficient given the magnitude of the transformative power of these technologies. As a species profoundly driven by what I called *Neuro P5 (power, profit, pleasure, pride and permanency),* it is crucial, going forward, to restrict the development of technologies that have the potential to lead to new forms of domination among social groups, or ultimately, of non-human entities over us.

Nayef Al-Rodhan

Prof. Nayef Al-Rodhan (@SustainHistory) is a Neuroscientist, Philosopher and Geostrategist. He is an Honorary Fellow at St Antony's College, University of Oxford, and Senior Fellow and Head of the Geopolitics and Global Futures Programme at the Geneva Centre for Security Policy, Geneva, Switzerland. Through many innovative books and articles, he has made significant conceptual contributions to the application of the field of neurophilosophy to human nature, history, contemporary geopolitics, international relations, cultural studies, future studies, and war and peace.