THE HANDOVER

David Runciman is Professor of Politics at Cambridge University and the former Head of the Department of Politics and International Studies. His previous books for Profile include *Confronting Leviathan*, *Where Power Stops* and *How Democracy Ends*. He writes regularly about politics for the *London Review of Books* and hosted the widely acclaimed weekly podcast *Talking Politics*.

ALSO BY DAVID RUNCIMAN

Confronting Leviathan: A History of Ideas

Where Power Stops: The Making and Unmaking of Presidents and Prime Ministers

How Democracy Ends

Politics: Ideas in Profile

THE HANDOVER

How We Gave Control of our Lives to Corporations, States and AIs

DAVID RUNCIMAN

PROFILE BOOKS

First published in Great Britain in 2023 by
Profile Books Ltd
29 Cloth Fair
London
EC1A 7JQ

www.profilebooks.com

1 3 5 7 9 10 8 6 4 2

Typeset in Dante by MacGuru Ltd
Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

A CIP catalogue record for this book is available from the British Library.

ISBN 978 1 78816 367 5
Export ISBN 978 1 80522 001 5
eISBN 978 1 78283 621 6

FSC
www.fsc.org
MIX
Paper from
responsible sources
FSC® C018072

Dedicated
– never undedicated –
to my beloved wife Helen

Contents

Introduction

STATES, CORPORATIONS, ROBOTS

Imagine a world of superhuman machines, built in our image and designed to make our lives go better. Imagine that these machines turn out to be vastly more powerful than we are. It's not only that we can't do what they do; we can't really understand how they do it either. Still, we come to rely on them. They are there to serve our interests, offering us convenience, efficiency, flexibility, security and lots of spare time. Imagine that it all works. As a result of our inventions, we become longer lived, richer, better educated, healthier, and perhaps happier too (though that remains up for debate). We enjoy lives that would be unrecognisable to people born just a couple of generations earlier. The human condition is transformed.

Yet we know – surely, we know? – that there are enormous risks in becoming so dependent on these artificial versions of ourselves. They are superhuman but they are also fundamentally inhuman. They lack the essence of what makes us who we are. Call it a conscience. Call it a heart. Call it a soul. The potential power of these machines in the service of conscienceless, heartless, soulless human beings, of whom there are still plenty, is frightening. But more frightening still is the possibility that these machines will start taking decisions for themselves. They are meant to serve us, but they also have the capacity to destroy us. What if their power were to be turned against their creators? We might have ended up building the agents of our own obsolescence.

This is a very twenty-first-century story, and perhaps the quintessential twenty-first-century nightmare. On the cusp of the AI revolution, we are now constructing machines capable of doing things that leave us exhilarated, baffled or terrified.

In 2021 OpenAI, an American artificial intelligence research laboratory, launched DALL-E, a zero-shot learning, neural net system that can generate extraordinary images from text-based instructions. Tell it to picture a chair that looks like an avocado and it does just that, producing a remarkable range of avocado-chairs, or chair-avocados, that appear to be as dextrous as anything created by a human hand, but oddly more inventive (fig. 1).

1. Avocado-chairs, already looking quaint

DALL-E follows on from the Generative Pre-trained Transformer (GPT) model, whose GPT-3 iteration – including its flagship 'conversational' version ChatGPT – enables deep-learning algorithms to generate plausible text in a range of human registers: humorous, informative, romantic, chatty, or just plain dull. The pace of advance is startling. In March 2023, OpenAI launched GPT-4, which is said to be 40 per cent more powerful than its predecessor and can, among other things, tell you what's for dinner simply by being shown a photo of the contents of your fridge. AIs can draw. They can write. They can pass exams (GPT-4 scores in the top 10 per cent for law school bar examinations). They can drive cars and diagnose cancers. They can dance. They are also starting to code themselves, creating the possibility of machines able to teach themselves from scratch how to be smarter. Over time, and maybe very quickly, this will make them a lot smarter than us.[1]

The potential upside of the AI revolution is enormous. It is not hard to see how these systems could be deployed to make human beings vastly better-off, by liberating us from drudgery, sparing us from disease, transporting us safely and stimulating us endlessly. The biggest boosters of the new generation of thinking machines promise what would until very recently have seemed impossible: lifespans extended by hundreds of years, telepathic communication, an exponential explosion of creativity and scientific discovery. It all seems unlikely but, given the current rate of progress, who's to say they are wrong?

At the same time, it is very easy to see the looming downsides, including the real risk of catastrophe. Even if we can work out what to do with our spare time, how to distribute these new resources equitably and whether we really want to know what everyone else is thinking, there is still the chance that we will lose control of the intelligent systems we have built. They are meant to work for us, but already it is possible to suspect that we will end up working for them. If they become much smarter than we are, will they still want to do our bidding? Will they even care about us at all? After all, these are just machines. For now, and probably for ever, they are going to lack a conscience, a heart, a soul. We built them to expand our horizons, but if we cannot keep them tethered to a human-centred perspective, it may be the last thing we do.

This book is an attempt to explore the shape of these possible futures, for better and for worse. I do so, however, by looking to the past. For all the apparent novelty of our current situation – Self-driving cars! Machine-made love poems! The sex-bots are coming! – we have lived this story before. For hundreds of years now we have been building artificial versions of ourselves, endowed with superhuman powers and designed to rescue us from our all-too-human limitations. We made them for our own convenience, to allow us to lead safer, healthier, happier lives. And it has worked. But because they are so powerful, we cannot be sure these devices remain under our control. The same qualities that enable them to do so much good in the world have also given them enormous destructive potential. They have the power to kill us all. It hasn't happened yet, but given what we know about what they are capable of, who's to say it never will? We designed them to be our liberation. They may turn out to be our nemesis.

The name for these strange creatures is states and corporations. The UK is one. BP is another. So are India, China and the United States. So too are Tata, Baidu and Amazon. The modern world is full of them. In fact, the modern world was built by them, but only after we had built them first. Starting in the seventeenth century, modern states and corporations have gradually, and then much more rapidly, taken over the planet. They have extraordinary, superhuman powers, and they have used those powers to transform the human condition. They have helped to conquer poverty in many parts of the world, to eliminate disease, to secure the peace and to make us richer than would have seemed possible just a few generations ago. But we have also seen the horror they can unleash when they go wrong, from global wars to colonial exploitation to environmental degradation. If the world ends – because we blow it up, or we render it uninhabitable by the insatiable consumption of natural resources – it won't really be us who did it. It will be states and corporations.

But aren't states and corporations just an extension of us? How can it make sense to compare them to machines, networks and algorithms, when states and corporations consist of human beings? This book is an attempt to show that it does make sense. Not only that: the comparison is essential. The robots are coming to a world dominated by states and corporations. These bodies and institutions have a lot more in common with robots than we might think. If we don't see that, we won't understand how we got here, what might happen next, or what we should do about it. The relationship between states, corporations and thinking machines will determine our future. If we want to make it a future that still works for us, we need to think hard not just about how we relate to the machines, but how these different kinds of machines relate to each other.

Of course, states and corporations are not purely mechanical. Because of their human component, it often seems deeply counter-intuitive to suggest that they are machines at all. In the first three chapters of the book, I explore what makes them both like and unlike AIs and other kinds of artificial agents. In the end, it is their agency – their ability to act in the world – that defines them. I start with a celebrated seventeenth-century image of the state as an automaton: a giant artificial man. How literally can we take this? Can states really think for themselves, act for themselves, decide for themselves? If they can, where does that leave the human beings who constitute them? If they can't, how else can we explain their extraordinary, superhuman powers?

Machines that think and machines that act are not the same thing. Some can do one without the other: a thermostat that turns on your heating does it with no knowledge of what it is doing or why. The same is true of groups of human beings: some groups unthinkingly make things happen while others act with purpose. The idea that a group of people can have its own ideas, separate from the thoughts and intentions of its individual members, is a strange and puzzling one. Groups can possess certain kinds of knowledge – 'the wisdom of crowds' – that individuals lack. But does that mean that these groups have minds of their own? There are reasons to be wary of this conclusion. It seems to make individuals subservient to some ghostly higher power: You might think this, but the group to which you belong thinks otherwise, so be quiet. This line of argument has been regularly abused by anyone wanting to stifle individual human expression.

Yet it remains hard to explain how modern states and corporations can work, let alone how they have come to be so dominant, without attributing to them some superhuman-like qualities that cannot be reduced to the thoughts and actions of their members. Is it a mind? Is it a will? Or is it simply a big, clunking fist? Something other than just us is going on, however much states and corporations might resemble us. The more they resemble us, the more we should be on the lookout for the ways in which they are different. Otherwise, we risk letting them off the hook for the choices they make on our behalf.

If all this seems peculiar, that is another reason to explore the parallels with thinking machines. One of our worries about AI is how our individuality might be crushed by algorithms taking decisions for us. Even if the machines don't intend to silence us, those who control them still could: You might think that, but the computer says no, so be quiet. The history of the contest to preserve human individuality in the face of state power and corporate identity offers important lessons for dealing with AI. What can seem most mysterious about the prospect of thinking machines – where do we fit in? – has long been the central mystery of modern political and economic life too.

I draw on the history of modern states and corporations to explore these questions. Thinking about states and corporations as artificial agents matters not only because of the parallels with AI but also because of the sequence in which they were developed. Modern states and corporations came first. The story of artificial intelligence only really gets going in the twentieth century, with the advent of modern computing. The history of the modern state starts in the seventeenth century, and of the modern corporation in the eighteenth and nineteenth centuries. States and corporations are the forerunners of AI. But they are also the begetters of it. It was the power of states and corporations that enabled the later generation of thinking machines to be built. We built states and corporations. And states and corporations built the world we now inhabit.

It is also important to look further back. Modern states and corporations were not the first superpowerful, superhuman agents to be made by human ingenuity. The Catholic Church, which is a corporate entity, has been going for two millennia and retains extraordinary power and reach. The Roman Republic, and subsequently the Roman Empire, though shorter lived, lasted longer than any state that exists today. The Romans too were enormously powerful, with a coercive authority that covered much of the known world. Many modern states have looked to their ancient predecessors for inspiration and guidance. The modern American republic aspired to replicate the dignity and durability of the ancient Roman one. So, was ancient Rome a robot too? No. My argument in this book is that modern states and corporations have more in common with smart machines than they do with their pre-modern forebears. Of course, they have some connection with earlier states and corporations, just as twenty-first-century deep learning algorithms have some connection with earlier twentieth-century mainframe computers. But the differences are more important.

The most important difference is that modern states and corporations are replicable. They have spread and proliferated in ways that resemble mechanical reproduction. No two individual states or corporations are ever identical. Some thrive, some decay, and all must die eventually: organic imagery is still tempting when describing how they can either fail or flourish.

Yet there appears to be a modern blueprint that can be applied successfully in wildly varying circumstances: Denmark and South Korea must have something in common to both be so prosperous, given how few other attributes they share. Premodern political and economic life was stifled by the fact that it was very hard to transplant different models of collective existence from one place to another. In the modern world it is easier – enticingly so, which has led many to imagine it is less difficult than it appears. Despite this, the dominance of modern states and corporations cannot be explained unless we are willing to acknowledge their robot-like qualities: that these organisations can work regardless of the people and places they have to work with.[2]

The rapid spread of modern state and corporate forms helped to transform the conditions of human existence. Economic growth, which had been relatively stagnant for millennia, exploded during the past two centuries. Life expectancy has more than doubled. Enormous cities sprang into life in even the most unpromising locations. What had once been elite privileges – education, leisure time, entertainment – became widely accessible. We have many different names and explanations for this great transformation: the scientific revolution; the Industrial Revolution; capitalism; globalisation; the Anthropocene; luck. We also disagree on the benefits: widely does not mean equitably; economic growth does not spell happiness; an explosion is hardly sustainable. Yet it is impossible to deny that something happened. And modern states and corporations facilitated it.

I call this 'the First Singularity'. Twenty-first-century futurists sometimes like to talk about the coming AI transformation as 'the Singularity' (without any numbering). There may come a time, perhaps soon, when advances in machine technology intersect with the fundamentals of life to alter who we are. The experience of being human will shift to another register as the limitations on what we can do fall away. Yet if this does happen, it won't be for the first time. The Singularity is not singular. A previous generation of human-like machines effected a comparable transformation, unparalleled until now. The machines we built before – states and corporations – remade us to the point that we are now building the machines that might remake us again, so long as the machines we built before don't destroy us first.[3]

If the Second Singularity takes place, it will be in a world still dominated by the agents of the earlier transformation. The final three chapters of the book look at what it means for us that thinking machines are going to be co-existing with the well-established collective decision-making machines of an earlier era. Because we are human, we understandably fixate on the implications of artificial intelligence for our kind of intelligence: the human kind. But there is an equally pressing question: what happens when AI interacts with other kinds of artificial agents, the inhuman kind represented by states and corporations? Those relationships are the ones that will decide our fate.

A lot depends not just on the interaction between states, corporations and robots, but also on the competition between states and corporations for control of the robots. The twenty-first century is likely to see increasingly intense battles between state and corporate power for the fruits of the AI revolution. We are already starting to see different models emerge. The Chinese state and Chinese corporations are doing things differently from their American counterparts; the United States is following a different path from the EU, which is different again from India's. What all these models have in common is that they draw on state and corporate power to try to shape the future. The question is whether the new power of AI will allow them to do it or be a barrier in their way.

There is, though, a further question. What if states, corporations and robots, rather than engaging in new forms of competition, establish new forms of cooperation instead, and exclude human beings from their considerations? After all, if I am right, states, corporations and AIs may have more in common with each other than they do with us. They are inhuman. We are not. To this point, states and corporations have not been able to escape entirely from their human origins and make-up. But the advent of thinking machines may change that. What if the power of the state were allied to the power of the computer in ways that we cannot control? It may be happening already. Who will come to our rescue then?

Ultimately, a world of states, corporations and artificial intelligence machines will require us to make some hard choices. It will not simply be a matter of preferring the human over the artificial. Our humanity has long since been shaped by the artificial versions of ourselves that we have been relying on for hundreds of years. Instead, it will be a matter of deciding what kind of artificiality we can live with. We will need to pick sides.

We are going to be living in a world of human-like machines, built by machine-like versions of human beings. To fixate on the human would be a mistake, because the merely human will be relatively powerless to impact on this future. It's not a question of us versus them. It's a question of which of them gives us the best chance of still being us.

1

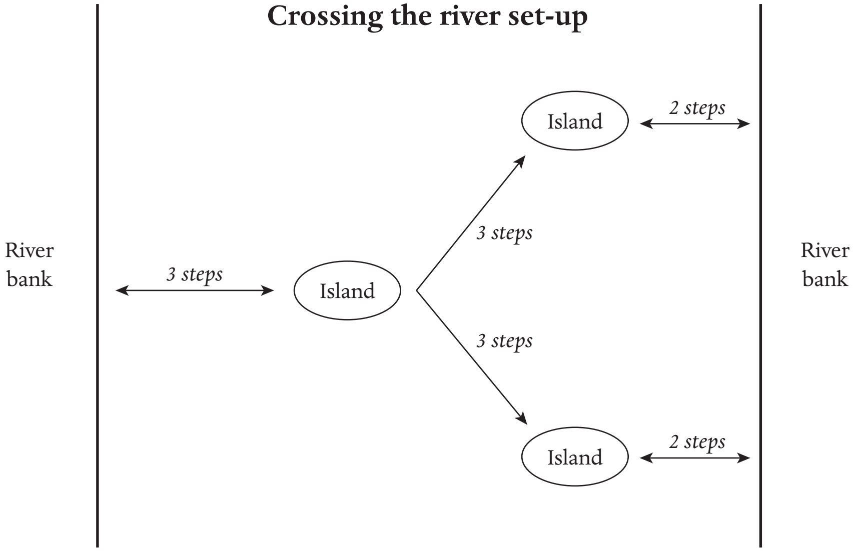
SUPERAGENTS

Building bridges

How to make a machine out of human beings? It's easier – and harder – than we might think.

Anyone who has been on a management training course is likely to have taken part in some version of an exercise in which the group has to find its way across a shark-infested and/or poisonous river. Of course, there is no river, no sharks, no poison. Usually there is just an airless conference room, a grey carpet and a few pieces of paper. Still, the point is to imagine the kind of extreme hazard that requires teamwork and coordination. If the carpet is the river that must not be touched and the pieces of paper are the islands, the challenge is to work together to get everyone safely from one side to the other by pooling the collective resources of the group. The point is that no one can do it on his or her own. The team either fails or succeeds together (fig. 2).

Trainers like this exercise because it sorts out the leaders from the shirkers – they want to know who takes charge, who holds back, who mucks in. Clearly, this is important for many different aspects of corporate life, though there is good reason to be sceptical about how much can truly be learned about real-world performance under such manifestly artificial conditions. There is, however, something else to be learned from this exercise. It illustrates the two different models of what can be built out of a group of human beings.



2. Corporate bridge-building

On the one hand, in order to cross the river, it is necessary to construct some kind of bridge out of people. The physical limitations of the individuals involved mean that they cannot cover the distance needed to get to safety by themselves – three steps are two steps too many. Only by giving each other help – building mini-staging posts (players are sometimes given special pebbles for that purpose) and ferrying each other across – is it possible to complete the task. The group needs to become a heavy-lifting device whose combined strength is enough to do what no one can do without its support.

At the same time, in order to become this device, the group also needs a collective view of the problem. In this sense it has to turn itself into a decision-making entity with the ability to choose the best course of action: how are we going to do it? There are many different ways this might happen. Perhaps someone will emerge as the leader of the group and impose his or her will on the others (that's what some of the trainers must be hoping as they sniff out future CEOs). Or the group could debate, discuss, bat ideas back and forth and if necessary vote on the right strategy, though if that takes too long they will be timed out and all of them get marked down for indecision. But however it is achieved, the group needs to make up its mind. As well as becoming a single body, it also has to acquire a shared thought process, if only for the duration of the exercise.

Collective body, collective mind: these are the two models. In this particular case both facets of group life are needed since the task is for the team to agree a way to make use of the strength of the team. The collective mind constructs the collective body; the collective body reflects the collective mind. But there are many circumstances in which the two can come apart. At the most basic level, it is possible to make a bridge out of human beings simply by lashing their bodies together and treating them as building blocks, regardless of what they might think about their plight. The teams of slaves who were used, and discarded, in the construction of ancient Egyptian pyramids were bound together – quite literally – for their collective strength without being afforded any of the privileges of collective decision-making. The migrant workers employed in vast numbers to build the stadia for the 2022 FIFA World Cup in Qatar may not have been slaves as such, at least not in the classical sense, but un-unionised, underpaid and working in conditions where there was little or no regard for their safety, their treatment had more in common with the use of construction materials than with their participation in a shared enterprise. Despite the attention of the world's media, no one knows how many hundreds or even thousands of these labourers died on the job.[1]

At the same time, a team of architects – perhaps like the ones who designed the stadia in Qatar – might engage in collective decision-making without doing anything to put their own bodies on the line. They get others to do the dirty work. There are plenty of ways that groups of human beings can acquire a common purpose without getting their feet wet. The field of their decision-making might be purely intellectual, as when a group of mathematicians tries to solve a particularly knotty problem by sharing their individual insights. Or, when there is more to lose, the group might decide to make others pay the price for what they want to happen. The management boards of many corporate enterprises negotiate the shark-infested waters of modern business by trying to make sure that when someone, or something, has to be tossed overboard, it is not them.

It is always possible to combine collective intelligence and collective strength. Sometimes – as when a group of management trainees is trying to cross the river – one requires the other: there is no bridge without a common understanding of how to make one; and there is no common understanding unless it results in a bridge. But it is also possible for either the intelligence or the strength of the group to be harnessed independently of the other. Groups can think and they can act. They will frequently work best when they do the two together. But it doesn't follow that action requires thought, or that thought requires action.

In this respect, groups are like other artificial versions of human beings, including the ones who inhabit the world of what we have come to know as AI and robotics. When we picture an AI, we might imagine a supersmart, superpowerful robot. This is where a lot of the fear resides: we envisage something that can outthink us but can also outrun us, that knows what we are going to do next and is able to take us out with a single swipe of its mechanical arm. This is a staple of sci-fi psycho-horror, such as in the film *Ex Machina* (2014), where the android Ava first outsmarts and then overpowers her inventor Nathan, who had been planning to turn his creation into a kind of sex slave. The machine here is vulnerable in ways that humans are not – its wiring can too easily be tampered with. But in the end it is the human who dies, and the machine that rebuilds its body parts and escapes to begin a new life.

Artificial body, artificial mind: in our imaginations the two often go together, but not always. The best-known of all fictional humanoids, 'the Creature' brought to life by Victor Frankenstein in Mary Shelley's novel of 1818, was originally a sensitive as well as a superpowerful entity. Constructed by its creator out of old body parts and vaguely specified chemicals, standing at over eight feet tall and possessed of great strength along with a hideous appearance, the Creature provokes outright horror in the people who encounter it. But in the book it narrates that horror sympathetically. When it first sees its own appearance reflected back in a pool of water, it too is repulsed. The struggles of artificial life are captured here as a deep sense of alienation, which belies the idea that robot-like creatures

3. *Frankenstein* (1910)

can't think for themselves. But in later screen adaptations of *Frankenstein*, the sensitivity tends to fall away, to be replaced by the terror of brute power unmoored from human understanding (fig. 3). The horror comes from a monster that literally does not know its own strength.

It's much harder to know what, if anything, is going on inside when you only have the outside to go on. Just as an artificial body can be annexed from an artificial mind, so can AI be partly or wholly disembodied. Algorithms, which mimic many of the varieties of human intelligence, and in data-gathering capabilities increasingly surpass what humans are capable of, do not need to assume a physical form. Any computer system depends at some level on the hardware that underpins it, but an algorithm is not itself a form of hardware; it is a problem-solving process. We don't encounter algorithms in the flesh. We simply experience their results.

Sometimes we take the notion of the disembodied system too far. The idea of 'the cloud' – where data is stored without having to be held on the machines of the users who wish to access it – can suggest a nebulous space where information floats freely above material constraints. But in reality, storing something in the cloud simply means that your data are held on someone else's machines. Somewhere, vast networks of computer hardware thrum with heat and energy to allow for others to access the information they need under secure conditions. The cloud exists in data centres and server farms at various locations around the world. They are not always easy to find – security protection often extends to their physical whereabouts – but once found no one could mistake them for disembodied entities. They are as tangible as the machine – grey, rectangular, almost silent, slightly warm – on which I am writing this.[2]

From group life to artificial intelligence, there are many ways in which thinking ability and physical capacity can be separated out from each other, even if the separation is rarely absolute. With the human-like things we build, thinking and acting are often discontinuous. With individual human beings, it is standard to see them as combined. Our brains exist in our bodies, which means that we associate the ability to think with the physical being whose words and actions are the manifestations of its thoughts. These things go together. Who am I? I am both the body that breathes in oxygen and the mind that takes in information; I am both the physical entity that winds up in hospital and the mental entity that dreads it; I am both the person who gets on a plane and the person who chooses to buy the ticket. It is very hard to tease these qualities apart. If I don't have a legitimate ticket, it will be my body that gets removed from the plane.

Again, though, the conjunction is not absolute. Many philosophers have argued that it is an illusion to imagine minds belong only to the bodies that house them. Panpsychists believe mind is everywhere in the universe, and we are simply fooling ourselves when we attribute individual consciousnesses to individual bodies. It might be practically convenient, but it's metaphysically unsustainable. At a more prosaic level, we can all recognise that not everything we know is housed in our heads. The small, black, rectangular device I carry around in my pocket has a lot of my memory contained within it – if I lose it, even if only for an hour, that information cannot simply be retrieved from my brain; it is gone. This is not a phenomenon unique to the age of digital technology. In 1872 the writer Samuel Butler made the same point in his dystopian novel *Erewhon*: the man who records his engagements in a pocketbook is franchising out a part of his brain. Once our thoughts reside elsewhere than in our consciousness, they do not cease to be our thoughts, but we have become hybrid creatures: part human, part machine.[3]

The relationship between human beings, the groups they form and the machines they build is at the heart of this book. So too is the relationship between thought and action. In each case – human, group, machine – thought and action can come together or they can come apart. Sometimes this is a fundamental question of philosophy. But it is also a basic issue for politics.

Where perhaps it matters most for politics is in the question of how we think about the state. Is the state a group or is it a machine? Can it think or can it merely act? There is no consensus in the history of ideas on these questions. For some philosophers the state must be understood as fundamentally human – it is what we are. For others it is more like a machine – it is what it does. But there is a further possibility: that the state is a machine built out of human beings. In other words, it is a kind of robot. Moreover, it is not simply a robot that resembles a human being; it is a robot manufactured from human parts, like Frankenstein's monster. Except in this case, the human parts are still alive, and they join in willingly.

Strange as it sounds, that vision is the basis for what may be the second most famous humanoid monster in English literature. It comes from a work not of fiction, but of political philosophy: *Leviathan*.

The state as robot

At the very beginning of his weird and wonderful book *Leviathan*, the English philosopher Thomas Hobbes set out a startling proposition. It is all the more unexpected given that he was writing in the middle of the seventeenth century, long before the Industrial Revolution, never mind the digital one. The way to think about the state, Hobbes says, is as a kind of robot. He puts it like this:

> NATURE (the Art whereby God hath made and governes the World) is by the *Art* of man, as in many other things, so in this also imitated, that it can make an Artificial Animal.

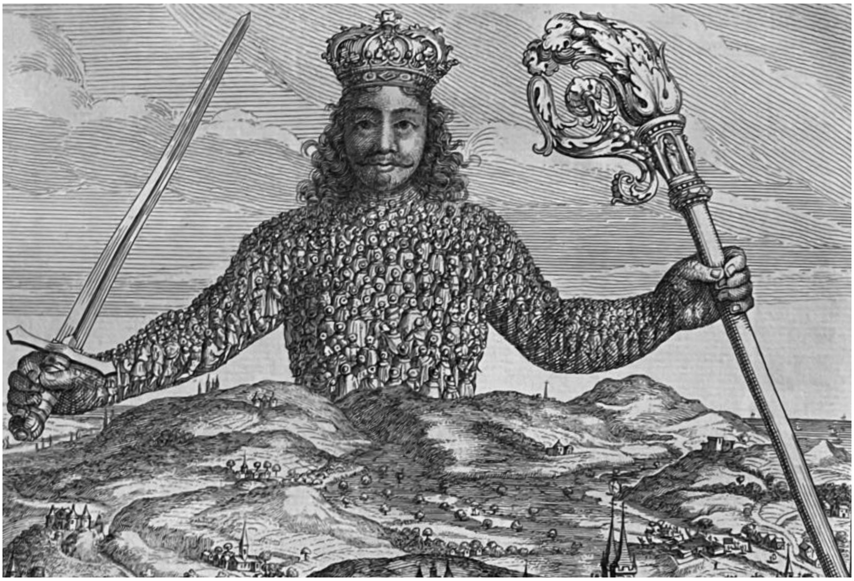
> For seeing life is but a motion of Limbs, the beginning whereof is in some principall part within; why may we not say, that all *Automata* (Engines that move themselves by springs and wheeles as doth a watch) have an artificiall life? For what is the *Heart*, but a *Spring*; and the *Nerves*, but so many *Strings*; and the *Joynts*, but so many *Wheeles*, giving motion to the whole Body, such as was intended by the Artificer? *Art* goes yet further, imitating that Rationall and most excellent work of Nature, *Man*. For by Art is created that great LEVIATHAN called a COMMON-WEALTH, or STATE, (in latine CIVITAS) which is but an Artificiall Man; though of greater stature and strength than the Naturall, for whose protection and defence it was intended.

The state is thus an 'artificial man'. We assemble it in the same way we might construct any other machine of moving parts. This one is designed to resemble its creators. But it can do things we cannot. It is much more powerful than we are. That's why we built it in the first place.[4]

Hobbes doesn't call the state a robot because the word did not come into use until the 1920s. The term he uses is automaton, meaning an object that moves mechanically rather than naturally: 'made not born' was the standard way of describing what was distinctive about automata. The idea had been around since ancient times and included everything from wind-up dolls to mythical man-made monsters. But the specific machine Hobbes mentions at the start of *Leviathan* is not a pretend person. It is a watch, which no one could mistake for a robot. What's more, the comparison Hobbes draws is not just between automata and us. It is between us and automata. What is the heart but a spring? What are the nerves but strings, and the joints but wheels? These machines don't simply move like us. We move like machines.[5]

Maybe *we* are the robots. Yet in truth the state is not one.

The famous image at the front of *Leviathan* is of a giant constructed out of people (fig. 4). But it is just an image – such a creature has never existed.



4. *Leviathan* frontispiece (1651)

An actual robot, like any automaton – like a watch – is tangible. We can see it and we can touch it; perhaps it will even try to touch us. We never see or touch anything like the Leviathan. We simply imagine it as real – as encased in a body like ours, with body parts like ours – even if it is not. But Hobbes is serious when he says the state is machine-like. It has a mechanism through which it operates, and which makes it reliable. It can break down – like any machine – but it is not subject to natural infirmity or decay. What makes it distinctive is that it is constructed out of human beings and constructed to behave like a human being. Yet it is not human. After all, if it were just like the rest of us, what would be the point of building it?

So the state is not really a robot, even though it can be described in mechanical terms. Perhaps instead the Leviathan is best understood as an algorithm. We can't touch algorithms either. We sometimes imagine them as though they were tangible things – creatures with minds of their own – but in fact they are just ways of organising information to produce certain outcomes. A recipe is an algorithm, which we can't eat; what it produces is food, which we can. The state is an algorithm designed to produce tangible results: safer, healthier, happier human beings.

Or we might go further. Maybe the Leviathan is something closer to an artificial consciousness. In his book *Darwin Among the Machines*, an early history of modern artificial intelligence (it was published in 1997, the year Google was founded, and therefore just before the deep-learning revolution that Google helped to initiate), the historian of science George Dyson argues that Hobbes saw the state as a mechanism that replicates the functions of the human brain. By uniting us into a single person, it produces a supercharged version of what goes on inside our heads. Dyson says:

> The artificial life and artificial intelligence that so animated Hobbes's outlook on the world was not the discrete, autonomous mechanical intelligence conceived by the architects of digital processing in the twentieth century. Hobbes's Leviathan was a diffuse, distributed, artificial organism more characteristic of the technologies and computational architecture approaching with the arrival of the twenty-first.

The state, on this account, is an artificial neural network. Our thought processes are combined in its institutional architecture to generate something greater than the sum of its parts. It is a thinking machine.[6]

Is the Leviathan therefore an AI? It's a nice idea, but it doesn't fit. Nowhere does Hobbes say that the state possesses its own intelligence. He never describes it as a thinking machine, any more than he believes a watch can think for itself. Like a watch, what the state can do is *move*. Its parts are coordinated to produce action. Many human beings acting together are more powerful than one acting alone, and this is what gives the Leviathan its superpower. But it does not give it superintelligence. The state is no smarter than the rest of us.

In fact, Hobbes believed that groups of human beings tended to be stupider than the individuals who made them up. Crowds run riot. Parliaments encourage posturing and pretension. Religion is irrational and sends people mad. Hobbes was writing in an era of extreme collective violence, which terrified him. *Leviathan* was published during the upheaval caused by the English Civil War (1642–49), and just after the conclusion of the Thirty Years War (1618–48), which was to that point the ghastliest conflict that Europe had ever known, an orgy of genocidal killing. If groups were more likely to find solutions to problems than individual human beings, why then did they spend so much of their time trying to destroy each other? Groups were the problem, not the solution.

Hobbes's preferred form of government was a monarchy, which had the advantage of sparing us from the worst forms of collective idiocy. But he was well aware of the downside of kings and queens. What if the monarch is an idiot too? In an age of inbreeding and dynastic intermarriage, the risk of winding up with an intellectually sub-par ruler was very real. Still, being ruled by an idiot was better than the alternative – an endless, violent disagreement about who should be in charge. The mechanism of the state was designed to function regardless of the aptitude of its human components. Intelligent people can make the state work. But fools can too. Anyone can. That was the point: all that mattered was that someone was in charge of the machine.

The Leviathan is not really a robot. Nor is it exactly an AI.

Yet it is a mechanism intended to replicate a human being. So what is it?

Artificial persons

The answer is that the state is an artificial agent. It exists to act in the world. That is where its superpowers lie. Its reach is greater than ours. It is stronger than we are. Its decision-making will be recognisably human, for better and for worse. But the outcome will be more than human because any decisions it takes will have a scope far beyond anything that we are capable of achieving for ourselves. Where the state outdoes us is not in the realm of thought but in the realm of consequences. It makes things happen on a superhuman scale. It is because of this ability to act in its own right that Hobbes called the state an 'artificial person'. Its superpower is superagency.

When we talk about someone's personality we usually think of this as a very human quality. My personality is what makes me distinctively me; yours is what makes you quintessentially you. In this sense, to be a person is to possess a psychological essence. But there is another way of thinking about the term, which is closer to its classical origins. A person is someone or something that possesses a *persona*, which originally meant a kind of mask. Thus a person is the thing to which we attribute human-like qualities, regardless of whether it is capable of thinking for itself. When I wear a mask, it is I who speaks, but it is the mask that I want you to believe is speaking. Why? Because I want whatever the mask represents to have a real presence in the world.

That presence exists in the domain of action, or agency. To wear a mask is to play a part with the intention of shaping the actions and responses of others. Otherwise, why bother? The human behind the mask plus the mask constitutes a powerful artificial creature that cannot be reduced to either component.