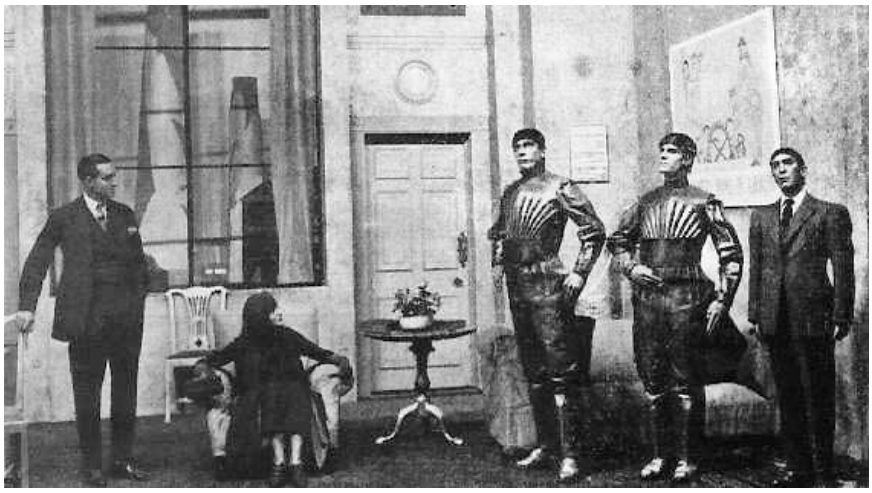
Minds, Machines, and Organisms: Robots in Western Philosophy and Culture

Gary Zabel

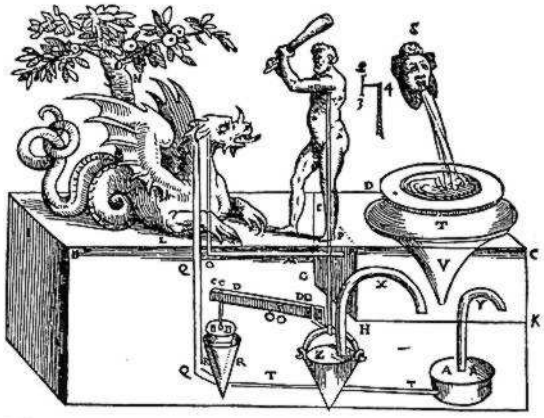


The word "robot" was invented in 1921 by the Czech writer Karel Capek as a name for the artificial humanoids who are the chief characters in his play, *R.U.R (Rossum's Universal Robots)*. Capek coined the name on the basis of the Czech word *robota,* which roughly means drudgery or servitude, a word appropriate for his dramatic purposes since the robots of his play are industrial slaves who mount a revolution against their human masters. Although Capek's robots were chemical creations, the word "robot" soon came to refer primarily to artificial creatures with mechanical bodies, sometimes, but not always, containing electronic components. As an expression for mechanical creatures, the word "robot" applies to a vast range of humanly produced objects going back as much as 2500 years prior to Capek's play. Before 1921, such objects were called "automata," derived from the Greek word *automatos*, which means having one's source of motion within oneself.

On the one hand, robots, or automata, are artifacts; in other words, they are products of human making. On the other hand, they share one of the principal characteristics of living things in that they are independently self-moving. In very odd and sometimes unsettling ways, they appear to bridge the gap between artifact and organism, the artificial and the natural, mechanism and nature. Because of their strangely intermediate character, robots have commanded the attention of poets, engineers, inventers, artists, biologists, and philosophers as well as the general public. In this course we will examine the role that robots have played in the history of Western philosophy and culture, focusing especially on the light they cast on the relationship between machines and living things. We will begin in ancient Greece where some of the earliest automata inspired the first efforts to develop a mechanistic theory of living things, as well as Aristotle's attempt to distinguish

between mechanism and the teleological, or goal-directed, character of organisms. We will then proceed to the Renaissance, where automata were regarded, at least symbolically, as magically animated artifacts, akin to living things. When these magical objects were incorporated, along with statues and fountains, into elaborate and beautiful gardens, some contemporary thinkers discerned the emergence of a "third nature," beyond the "first nature" of organic life and the "second nature" created by human technique. We will then leave the magic of the Renaissance for the mechanistic materialism of the Enlightenment. During that period, some thinkers began to see automata as models of the human body, and to regard human as well as non-human animals as robots of a very complex kind. In our studies, the Enlightenment will give way to the nineteenth century, in which the industrial revolution leads to the proliferation of automata of many kinds, including humanoid robots, as well as designs for an early programmable computer in the form of Charles Babbage's Analytical Engine. Though Babbage's plans to automate human thinking were more advanced than the technological capabilities of his own period, they came to fruition in the computer revolution of the following century. Leaving the 19th century, then, we will examine 20th century attempts to create mental automata, artificial minds capable of human-like thought. In this part of the course, we will examine the philosophical debate over artificial intelligence and its significance for our understanding of the human mind. Finally, we will investigate recent attempts to lodge artificial intelligence in robot bodies. In this context, we will examine the tension between deliberative and bio-robotics, as rival attempts to create thinking and living machines. This will bring us full circle to the beginning of the course, and the project on the part of ancient philosophers to fathom the relationship between mechanism and life.

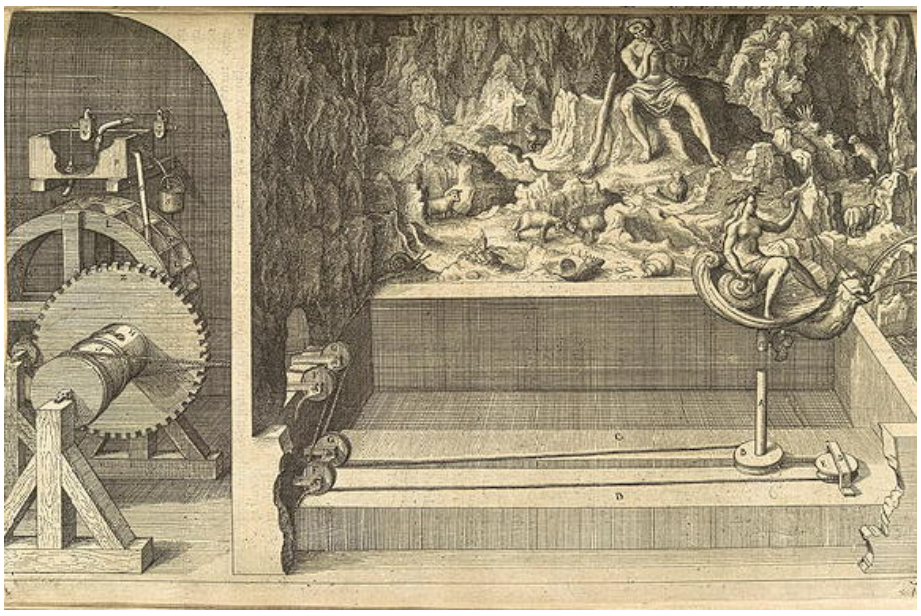
Ancient Robots and the Explanation of Living Things



Western philosophy originated in Greece in the 6th Century B.C.E. with the attempt of a handful of thinkers to replace mythical explanations of the origins and structure of the cosmos with natural ones. The tradition of naturalistic explanation that began with such philosophers as Thales, Anaximander, and Anaximenes achieved an especially advanced expression in the 5th Century B.C.E. in the atomism of Leucippus and Democritus. We will begin this section of the course by considering Democritus' version of atomism,

especially his attempt to explain life and mind by appealing to the entirely materialist picture of a world of indivisible particles interacting with one another randomly in empty space. We will then investigate Aristotle's critique of Democritean atomism as an approach to biological explanation. According to Aristotle, the explanation of living things by appeal to the principles of matter in motion is necessary but incomplete. Purely materialist accounts of biological organisms ignore their goal-directed, or teleological, character. Such teleology is present in both organic structure and process. It is expressed in the way the parts of an organism contribute to the meaningful pattern that shapes that organism as a whole, as well as the way the developmental stages through which with an organism passes occur for the sake of the fully developed adult that is in the process of emergence. However, it is not just that materialist explanation must be supplemented by teleological explanation in the science of living things. For Aristotle, the behavior of matter is a subordinate and instrumental element in the life of the organism; it serves the purposes of the emergence, development, and flourishing of meaningful organic form. In several passages of his biological writings, Aristotle refers, not merely to materialist accounts of living things, but to full-blown mechanistic ones. That is to say, he entertains, and even endorses, the idea that organisms can be understood, though only partially of course, on analogy with humanly produced machines, especially with automata, self-moving artifacts made in the image of living beings. The mathematician and engineer, Hero of Alexandria, wrote his treatise *Pnuematica* almost 400 years after Aristotle's death. However, some of the water-powered automata described in the treatise existed, perhaps in less developed form, even in Aristotle's day. We will read the *Pneumatica* for its accounts of the hydraulic robots of ancient Greece.

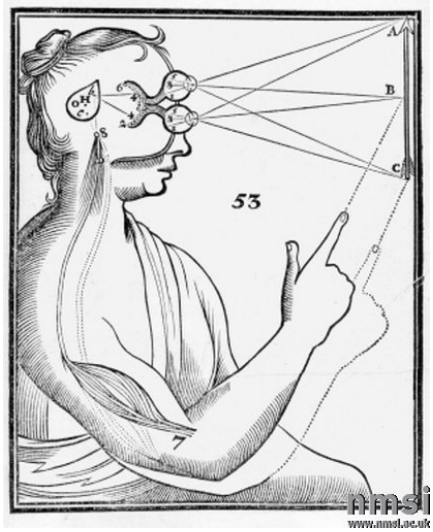
Renaissance: Robots, Magic, and Mechanism



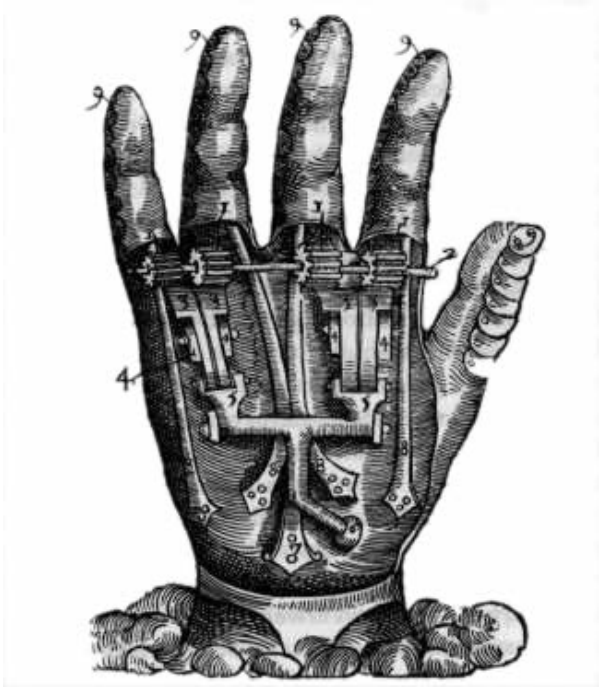
As we have seen in the previous section, automata have inspired tough-minded mechanistic explanations of living things. Our ability to produce life-like machines has

sometimes been taken as evidence that life itself is mechanical in character. However, automata also instill in many people a sense of the uncanny. Even in our age of computerized electronic robots, a purely mechanical automaton preserved, say, from the 19th Century can, to put it bluntly, give us the creeps. Automata have always been surrounded by an aura of the occult, as it has been a perennial dream of the magician to transform inanimate objects into living things. Nowhere has the connection between automata and magic been more pronounced than in the Italian Renaissance. The Renaissance involved a revival of ancient Greek tradition, including strains of learned occultism that had developed in the Hellenized world presided over by the Roman Empire. In particular, the rediscovery of the *Corpus Hermeticum* – a series of occult gnostic texts composed few centuries after the birth of Christ – may have stimulated a magical interpretation of automata. The *Aesclepius*, one of the texts of the *Corpus Hermeticum*, refers to an ancient Egyptian practice of animating statues by inducing gods to inhabit them. That passage inspired some of the writings of the influential Renaissance philosopher, Marcilio Ficino. It is hard to say whether Ficino's writings in turn inspired Leonardo Da Vinci, but plans for two remarkable automata – one of them apparently a robot guided by a programmable analog computer – do survive in Leonardo's notebooks. Whatever the relationship between Leonardo and Ficino, the latter's interest in magical animation left an unmistakable imprint on the use of automata in the astonishing gardens of the late Renaissance. Integrated with statues based on ancient Greek mythology, elaborate fountains, and extensive hydraulic machinery, Renaissance automata contributed to the emergence of what some thinkers saw as a "third nature" that both synthesized and surpassed the "first nature" of biological life and the "second nature" of conventional artifacts. In this section of the course, we will examine these themes by reading passages from the *Aesclepius*, a contemporary attempt to reconstruct Leonardo's robots, and a treatment of the automata incorporated into Renaissance gardens.

Enlightenment: The Human Being as Robot



Like the Renaissance, the European Enlightenment of the 18th Century represented a return to the intellectual resources of the ancient world. However, the Enlightenment thinkers rejected those strains of Hellenistic occultism that so fascinated Renaissance intellectuals, embracing instead Democritean atomism and other forms of ancient materialist and mechanist thought. For the key figures of the Enlightenment, the human body, like the bodies of other living things, is a machine. It can be understood reductively, by analyzing it into its parts, discovering the laws of nature that govern the motion of those parts, and, on the basis of such understanding, tracing the interactions of the moving parts over the course of time. There is no place for Aristotlean "final causes" in the explanation of nature. Teleology is swept into the dustbin of the history of science. In this section of the course, we will examine two versions of Enlightenment mechanism as applied to human beings. Descartes, who is usually regarded as the founder of modern philosophy, is a moderate mechanist. In his *Treatise on Man* as well as other writings, he draws on the medical knowledge of his day to provide an account of the human body as a complex machine, explicitly identifying it as an automaton in more than one passage. However, he also argues that the mind is not open to mechanistic explanation, because it is not extended in space, and therefore is not a material entity, let alone a machine. Cartesian dualism, of course, leaves the problem of understanding how mind and body interact. How is it possible for the immaterial mind to operate the mechanical levers of the body-automaton? The second Enlightenment figure whom we will consider resolves this problem by radicalizing Descartes' mechanistic science of the human body. For Julien Offray de Le Metrie, there is no ghost in the machine. Human beings simply *are* their bodies, and their bodies are mechanical automata. By rejecting Cartesian dualism, Le Metrie levels the ontological ground. There is only one plane of existence, and that is the natural world that serves as the theme of the science of matter in motion. We are a part of that world, no different in substance than any other part. People are robots.



19th Century Automata

The Enlightenment project for the mechanization of nature bore fruit in the 19th Century with the definitive triumph of the industrial revolution, first in England and the United States, and then in much of the rest of the Western world. The use of machinery in industry was nothing new. Machinery had been used in production processes, especially in mining, in ancient Greece and Rome, and the high Middle Ages (from around 1100 to 1400 B.C.E.) saw the flowering of new forms of machinery in farming, construction, sanitation, and so on. As Marx points out in one of our readings, what was unique about the industrial revolution stimulated by modern capitalism was that fact that multiple machines of distinct and varying characters were integrated into a whole system of mechanized production. In that system, human beings came to play a less and less central role. At first workers were incorporated into the machine system as extra bits of machinery, as the work was broken down into disjoint part-processes whose rhythm was determined by the pace of the machine system as a whole. But the reduction of human activity to subjectively meaningless repetitive mechanism was merely a prelude to the expulsion of workers from direct involvement in machine industry, since mechanized human activity is handled much more efficiently by real machines. As Marx points out, the mechanized factory is transformed into an automaton, and human beings take on the role of overseeing and tending these artificial self-moving creatures. Given the development of what we now call "automated" production, it is not surprising that the late 18th and early 19th centuries were also gripped by a fascination with androids and other mechanical "organisms." France in particular was the home of such craftsmen as Pierre Jaquet-Droz, Jacques de Vaucanson, and the Maillardet Brothers, whom we can only regard as the greatest creators of purely mechanical automata in Western history. Nineteenth Century audiences were entertained at fairs, world exhibitions, and stage shows by the animated ducks, elephants, writers, clowns, dancers, magicians, and musicians created by such master mechanics. While Jaquet-Droz and the others were

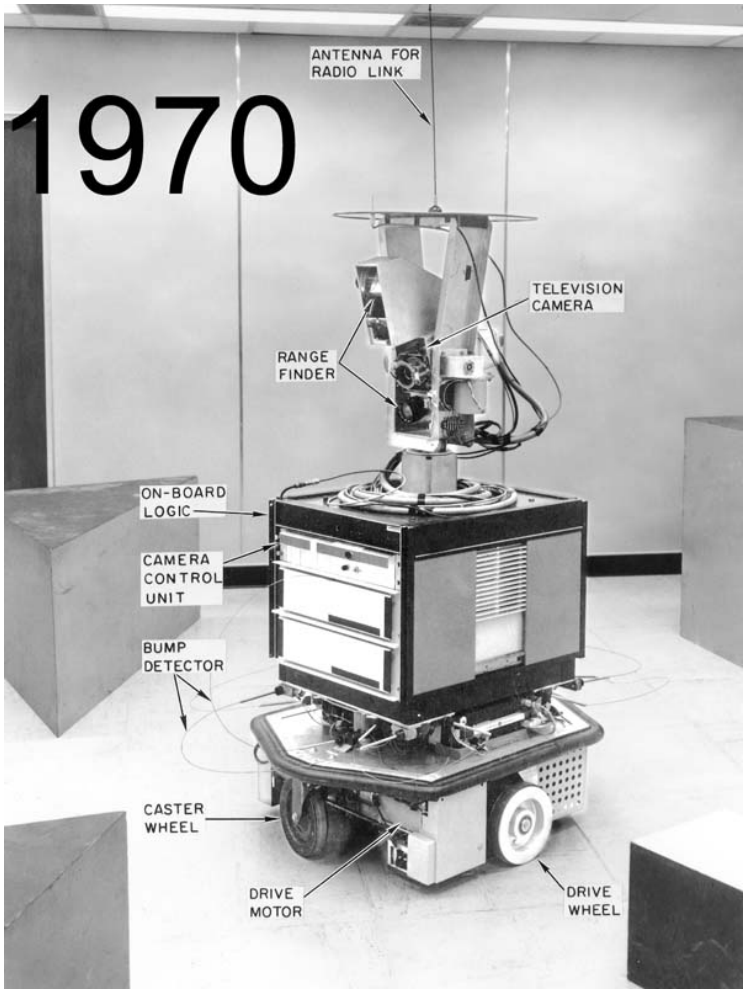
creating automated human bodies, the English mathematician and inventor, Charles Babbage, developed plans for machinery capable of automating the human mind. Though technological and financial limitations prevented him from building a full-scale prototype of his "analytical engine," we now recognize that his plans for that machine were perfectly viable. Babbage was the inventor of the memory-and-processor architecture that is still used in our computers, even though his machine had no electronic elements, consisting instead of the kind of elaborate gearing that drove the more conventional automata of the period. A reading of Marx and Babbage as well as images and film clips of late 18th and early 19th century automata will help us pursue these themes in this section of the course.

Artificial Minds

Emergence of the electronic programmable computer in the aftermath of the Second World War raised Babbage's plans for the mechanization of human thinking to the level of technological practicality. Alan Turing, who is often credited with inventing the modern computer, established in a famous paper a test for human-level intelligence and suggested that a digital computer would be able to pass that test by the end of the Twentieth Century. Though no computer, digital or otherwise, has been able to pass the "Turing Test" as of yet, Turing's work gave birth to the field of Artificial Intelligence,

which is alive and kicking half a century after his death. In its heroic early phase, the field of AI shared Turing's optimism about the immanent achievability of machine intelligence on a human scale. The theoretical framework that sustained that optimism is sometimes called Good Old-Fashioned AI (GOFAI). GOFAI operates with the so-called "physical symbol system hypothesis," which regards thinking as the rule-governed manipulation of definite and distinguishable physical marks, or "symbols." The rules of symbol manipulation are purely syntactic; in other words they are rules governing the combination of primitive symbols into properly formed complex expressions, similar to the way in which the rules of English spelling and grammar allow us to combine the letters of our alphabet into words, our words into sentences, and our sentences into paragraphs. The key characteristic of syntactic rules is that they are purely formal, and so are not sensitive to the meaning of words, i.e. to their semantics. The hope of the physical symbol system hypothesis is that the right set of syntactic rules will preserve meaning, in the way in which the purely formal procedures of deductive logic derive meaningful conclusions from meaningful premises, without those procedures themselves being sensitive to what the symbols they operate with mean. It is precisely that hope of the physical symbol system hypothesis that the philosopher John Searle challenges in his famous thought experiment of The Chinese Room. Searle imagines a person who does not speak Chinese inside a room with a slot through which she exchanges written Chinese expressions with a Chinese speaker outside of the room. The person in the room receives a card through the slot with a series of Chinese characters written on it. She consults a book which correlates those marks with a different set of Chinese characters. She writes down the new set of characters on another card, and passes it back through the slot. As the process continues, the person in the room conducts an entire conversation in Chinese with the Chinese speaker outside of the room, and yet she does not understand a word of Chinese herself. Now here is the point. Searle claims that the room is a digital computer, the person inside the room acting as its processor, and the book correlating characters as both its program and its memory. On the basis of this thought experiment, Searle argues that digital computers do not understand the expressions they manipulate any more than the person in the Chinese Room understands Chinese. Since understanding the meaning of expressions is a centrally important aspect of human thinking, no machine could possibly think just by performing digital computations. Proponents of artificial intelligence have crafted several rejoinders to Searle's argument, but the one that we will be concerned with in this section of the course was dubbed by Searle himself the Robot Reply. It is true, so the rejoinder goes, that the Chinese Room demonstrates that a conventional digital computer is unable to understand the meaning of symbols, but that is because its symbol processing does not have the right causal connections with the outside world. In order for the symbols to become meaningful, the program that encodes them would have to be run on a computer guiding the activities of a robot. The solution to what is sometimes called the "symbol grounding problem" is to allow the symbols to mediate the robot's interpretation of the data received by its sensors, and to steer its motor responses to such data. In this section of the course we will examine the Robot Reply as well as Searle's response to it.

Deliberative Robots, Biorobots, and Problem of Embodiment



Early attempts to create robots run by digital computers shared the emphasis on reasoning, or deliberation, that characterized AI research in general at the time. A classical example of deliberative robotics was the robot Shakey who roamed the labs of the Stanford Research Institute in the late 1960s and early 1970s. Shakey executed a program called STRIPS that generated a model of the environment the robot inhabited as well as logically coherent plans that enabled it to perform complex tasks in that environment. Like all purely deliberative robots, however, Shakey had certain rather severe problems. It took a great deal of time for the robot to generate plans, and it was incapable of adjusting to environmental circumstances that had changed after the completion of the planning process. Because of these difficulties, Rodney Brooks, then director of MIT's Mobile Robots Lab, advocated that the deliberative approach to robotics be replaced with a "behavioral" one. Behavioral robotics begins with the claim that the world is too complex to be modeled adequately in programming language. Instead of designing a robot that creates a world model and then generates logical plans to move within it, it is better to equip the robot with the sensors and simple interpretative programs that enable it to react immediately to the objects it encounters in the course of its activities. On the basis of this approach, Brooks built remarkable insect-like robots

that were able to achieve such simple goals as seeking out light sources while avoiding or climbing over obstacles. The reaction against deliberative robotics was expressed in a more radical form in the work of the Los Alamos roboticist, Mark Tilden. Brooks' robots were not logical planning machines, but they did depend upon digital processing to link sensory inputs and motor responses in what Brooks called "subsumption hierarchies," ordered patterns in which certain behaviors took priority over others. Tilden, however, argued that it was a mistake to rely upon digital computers to build competent artificial organisms. The brains and peripheral nervous systems of animals, so he claimed, are analog, not digital in character. By using inexpensive analog electronic components, Tilden has been able to build hundreds of robots that adapt their behavior to the surrounding world in flexible and subtle ways, unmistakably suggesting the illusive style of organic life. As a result of the work of Brooks, Tilden, and other researchers, "bio-robotics" – the attempt to model robots on the example of living organisms – is now a very lively field. In this section of the course, we will use the work of the philosopher Maurice Merleau-Ponty and the animal psychologist Jacob Von Uexkull to evaluate the bio-robotic response to the problems of logical AI. We will address ourselves two critical questions: are the bodies of electronic robots really capable of some version of life? and are the higher-order processes of logical thought merely sophisticated versions of organic response, or does the rational mind have an independent character?