6
MECHANISMS AND THE MENTAL

Marcin Miłkowski

1. Introduction: the notion of mechanism evolves with the understanding of the mental

In this chapter, I sketch the history of mechanistic models of the mental, as related to the technological project of trying to build mechanical minds, and discuss the changing use of these models. In section 2, I introduce the Cartesian notion of mechanism, which shaped the debate in the centuries that followed. Early mechanistic proposals are also connected with early attempts to formulate the computational account of thinking and reasoning, upheld notably by Hobbes and Leibniz. In section 3, associationist and behaviorist models of the mind are sketched, along with attempts to understand the neural system in terms of connections and associations. Early robotic models, built mostly by behaviorists and other students of animal behavior, are also introduced. In section 4, the focus is on early computational and cybernetic models of the mind. In section 5, I deal with computational models of mental mechanisms as proposed by students of artificial intelligence and cognitive science. The history of uses of mechanistic models sheds light on different kinds of explanations of the mental.

The notion of mechanism has played an immense role in the history of attempts to scientifically explain mental phenomena. On the one hand, it has always been related to the philosophical mind–body problem, and, on the other hand, to the possible scope of mechanistic explanations. Mechanization of thinking is also one of the classical problems in philosophy of mathematics, and the notion of mechanism, in a much more abstract form, appears in the history of mathematics as well.

The philosophical mind–body problem is the question of how the mental and the physical are related. One way to defend a naturalistic solution to the mind–body problem is to build a mechanistic model of the mind. Such a model might show, even as a proof of an in-principle possibility, that thinking can be instantiated in a physical mechanism. Conversely, the proof that thinking may not be mechanized hints at the possibility that the mind is not physical, after all.

However, one may also contend that mechanistic models of thinking, even if successful, do not decide the mind–body question at all. In other words, even if thinking is just the exercise of computational mechanisms, there still could be some non-mechanistic aspects of thought or perception that completely evade the mechanistic explanation. Indeed, this is the position taken by computational dualists, from Leibniz to David Chalmers. Yet the existence of limits to

mechanization may be interpreted not as evidence that human minds transcend mechanism but that they are also subject to the same limits, as the debate over limiting theorems in mathematics shows (see Chapter 33).

In short, the development of technological mechanisms has affected the ways one conceived of possible explanations of the mental, and mechanists tried to describe, build, and elucidate mechanisms that exhibit at least some cognitive capacities, including mathematical reasoning and conceptual thought. Early mechanistic models were just proofs of possibility of mechanization; however, current mechanistic explanations strive to go beyond the question of how it is possible to display such-and-such a mental phenomenon, and (at least partially) explain actual mental processes.

2. Early mechanists: Descartes and Hobbes to Leibniz and mechanical ducks

René Descartes is known for his defense of mind–body dualism, but he was also a staunch proponent of mechanistic explanations, which paved the way for modern science. His understanding of mechanisms was strongly influenced by the technology available in his age, in particular fountains and clocks. A *mechanism* for Descartes is a spatial system whose operation depends on local interactions among its parts. These interactions can be further described geometrically. Notably, mechanisms can be explained without recourse to teleology, and indeed, the Cartesian account of the mechanistic explanation is strongly opposed to the Aristotelian and medieval models of explanation, which used to appeal to final causes, theological considerations, angelology, basic elements, and Pythagorean principles, all at the same time.

The significance of Descartes does not stop here. He also offered early mechanistic accounts of the nervous system and psychological phenomena (see Chapter 28). The nervous system was conceived of in hydraulic terms of flows of animal spirits. These are produced in the blood and may exert influence over bodily parts. The basic building block of the nervous system, according to Descartes, is the *reflex arc*: the stimulus pulls tiny wires of the nervous system, which in turn open little valves in the brain, releasing animal spirits to hollow nerve tubes that lead to appropriate muscles (see Figure 6.1). This general outline of the reflex arc as the basic operation of the nervous system has remained immensely important, even if subsequent research rejected hydraulic metaphors. The modern notion of *reflex* itself was introduced by a medical doctor Thomas Willis in his treaty *De motu musculari* (Willis 1694). Willis was influenced by Descartes, but his mechanism was framed in optical terms (reflection being an optical phenomenon); the reflex is a pure reaction to external influence rather than something that relies on internal resources (as in Descartes' model (cf. Canguilhem 1955)).

However, Descartes argued that there are limits to mechanistic explanation of the mind. He thought that the physiological machine just described can have functions such as

> the reception by the external sense organs of light, sounds, smells, tastes, heat and other such qualities, the imprinting of the ideas of these qualities in the organ of the "common" sense and the imagination, the retention or stamping of these ideas in the memory, the internal movements of the appetites and passions, and finally the external movements of all the limbs.
>
> *(Descartes 1664/1985, 1: 100)*

But in *Discourse on the Method*, he claimed that no mechanism can be built that could "use words, or put together other signs, as we do in order to declare our thoughts to others" (Descartes 1637/1985, 1: 140). According to him, it is not conceivable that "a machine should produce

Figure 6.1 The reflex arc of the painful stimulus according to Descartes. The fire A is a stimulus afflicting the skin (B) and moving the fine thread (C), which goes to the valve (d, e). The valve opens the cavity (F), from which the animal spirit is released, and this in turn makes the head turn and move the hand and the foot

different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence." Moreover, no machine could work for a variety of purposes, and, in contrast, reason is a universal instrument. A human being's rational capacities must therefore come from a non-physical rational soul, which Descartes believed to be connected with the material brain through the pineal gland (Descartes 1649/1985, 1: 340). The Cartesian challenge is still alive today (Wheeler 2008).

In contrast to Descartes, Hobbes was probably the first proponent of a thoroughly mechanistic and computational theory of mind; as he claims, ratiocination is just computation: "to compute is, either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another" (Hobbes 1655/1839, 3). However, Hobbes failed to deliver even a sketchy description of a mechanism that could indeed think by computing. But soon Leibniz envisaged that any philosophical argument could be simply settled by computing in a universal mathematical language (Leibniz 1666). The seventeenth century witnessed important developments in the construction of various machines, including Blaise Pascal's arithmetic machine, but none of them was fit to perform complex logical computations. Further development in engineering led to even more sophisticated machines, the Digesting Duck, created in 1739 by Jacques de Vaucanson, being probably the most prominent. The mechanical duck appeared to be able to digest food; however, it only collected food in one container and produced artificial feces from another one. Another imaginary

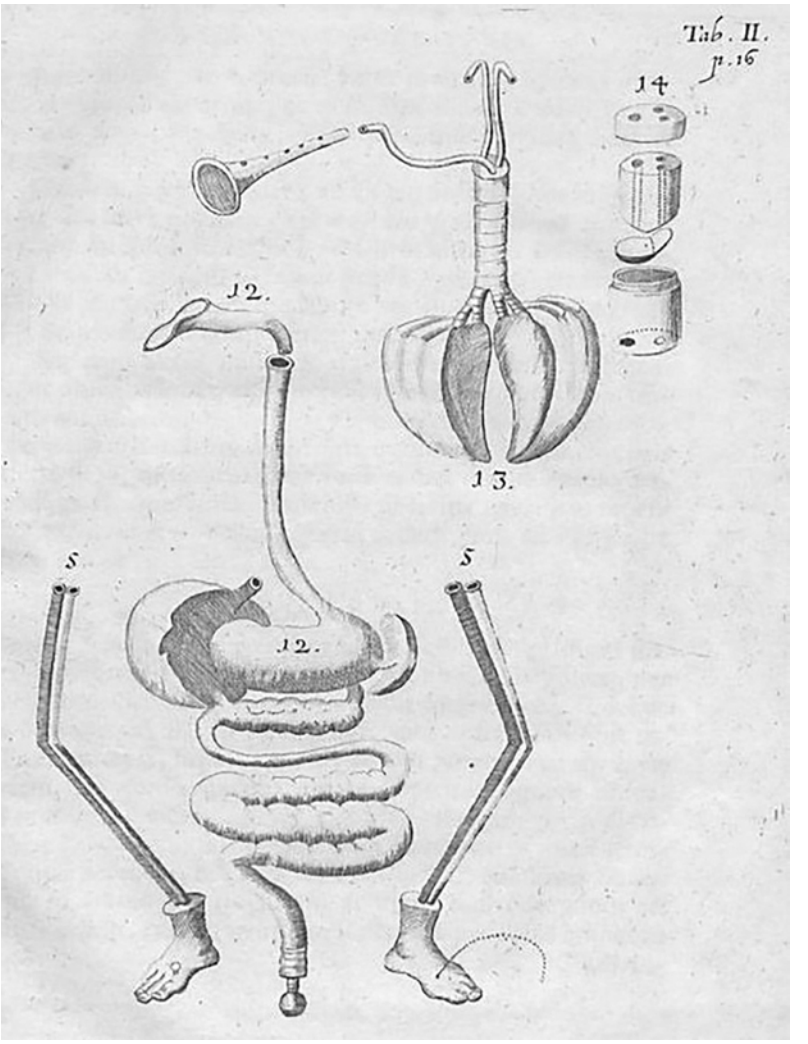

Figure 6.2 *Statua Humana Circulatoria*, Salomon Reisel's imaginary model of respiration

model of human respiration, *Statua Humana Circulatoria*, conceived by Salomon Reisel, is depicted in Figure 6.2. At the same time, such contraptions aroused the imagination of proponents of mechanical reductionism.

The mechanical reductionism is mostly associated with Julien de La Mettrie, who offered a purely mechanical account of the mental in his *Man a Machine* (La Mettrie 1748/1912). Note, however, that La Mettrie did not espouse the Cartesian notion of the mechanism; he claimed that the matter is capable of movement and sensibility only when appropriately *organized*. To him, organization is not just a matter of spatial arrangement, as in Descartes, but also of inner complexity that may be quantified. In this, La Mettrie embraced a notion of mechanism closer to the one defended by New Mechanists than the Cartesian one. At the same time, he admitted that it was an open question what kind of organization contributed to sensation.

3. Associationistic models of the neural system and early robotic models

The most influential criticism of the Cartesian account of thought came from John Locke, who criticized the theory of innate ideas defended by Descartes, and claimed that all thought came from the senses. Yet for the development of the mechanistic understanding of the mental, another empiricist claim is much more important: namely, that ideas form thoughts by association. The associationist theory of ideas has roots as deep as in Plato's *Phaedo*, and early formulations can also be found in Hobbes. But it was David Hume, having identified three principles of association—similarity, contiguity of time and place, and cause and effect—who had the greatest influence on subsequent theorizing about mental mechanisms. For example, having experienced similar impressions when touching fire, a stove, and hot metal, we may associate all of them with the notion of *something hot*.

Soon after, associationism was linked together with hydraulic models of the mind. For example, Sigmund Freud's speculations about neurological mechanisms underlying psychological phenomena relied on associationist and hydraulic hypotheses (compare Bilder and LeFever 1998). In particular, a difficult question was how new associations were formed. Associationism suggested that there had to be a certain mechanism, and as late as at the turn of the twentieth century, physical models of association learning were offered (for a comprehensive review of such models, see Cordeschi 2002).

In the second half of the nineteenth century, the first labs in experimental psychology were formed by Wilhelm Wundt in Germany and William James in the US. Wundt's work had deep roots in psychophysical research, whose champions included physicists Herman Helmholtz (who also posited unconscious inference) and Ernst Mach. However, Wundt's goal was to explain psychological phenomena with lawlike generalizations, just like in psychophysics, rather than by appeal to organized mechanisms. His experimental paradigm, related largely to introspection, did not uncover a large body of well-confirmed laws. In opposition to Wundt's introspective psychology and to its American competition in the work of Titchener, the behaviorist movement started. Interestingly, behaviorists, such as E. Thorndike, also tried to discover general laws of learning that would hold true of all animals.

Behaviorism is committed to a broadly associationist view of the learning organism. Hence, at the turn of the twentieth century, there was a large interest in understanding the underlying principles of association. This was also driven by new discoveries in neurobiology; in particular, Sir Charles Sherrington proposed that elementary building blocks of the nervous systems, the neurons, communicate via synapses (see Chapter 28). At the time, Santiago Ramón y Cajal and Camillo Golgi debated over the general anatomy of the nervous tissue. Cajal argued that the nervous tissue is composed of multiple cells while Golgi claimed that the brain is a single continuous network, called the *reticulum*. Sherrington's proposal of the synaptic communication, related to his detailed study of reflexes, was one of the tipping points in the debate (Burke 2007). Moreover, he also refined the Cartesian notion of the reflex arc as the basic operational blueprint of the nervous system. The notion of the reflex arc—as composed of a receptor, conductor, and effector—was compatible with the behaviorist psychology, which focused on stimulus-response pairs.

Such is the background for the early connectionist theory defended by Edward Thorndike (1911). According to him, "the animal mind is, by any definition, something intimately associated with his connection system or means of binding various physical activities to various physical impressions" (Thorndike 1911, 16). The "connection system" underlies the capacity of animals to learn and consists of connections between stimuli coming from the external environment and collected by a "receptor system," and motor responses.

Subsequently, artificial models of learning mechanisms were proposed. For example, Bent Russell built hydraulic networks that demonstrated effects of learning over time, supported by experimental results of reinforcement on animals (Russell 1913). This approach was then continued by a prominent behaviorist, Clark Hull, in his robotic models of learning in the 1920s and 1930s. Hull wanted to build robots to prove the possibility that complex forms of behavior were not limited to living organisms. Hull's models were no longer hydraulic but electrochemical. His model of the conditioned reflex was supposed to free "the science of complex adaptive mammalian behavior from the mysticism whichever haunts it" (Krueger and Hull 1931, 267). Importantly, it was not an exact replica of animal behavior but a *conceptual model* that proved that a number of features of behavior are mechanically replicable. This feature of simulative models remained important throughout the rest of the twentieth century.

However, the time was not yet ripe for simulative research and later Hull shied away from further robotic simulations, as they were not treated seriously. For Edward Tolman, with his hypothetical robotic "schematic sowbug," a simplified artificial animal that illustrated his theory that all behavior was entirely stimulus driven (Tolman 1939), robotic models were not essential in studying behavior either. Not only behaviorists built robots at the time; one particularly important exception was Alfred Lotka (1925), known for his work on mathematical predator/prey models. His simple mechanical animal model—a "correlating apparatus"—was an important predecessor of later cybernetic robots for two reasons. First, Lotka approached the evolution of complex systems in terms of reciprocal equilibrium. No longer does the nervous system work merely as a reflex arc; correlating actions with the environment is "essentially cyclic in character: It has its origin in the external world, which becomes depicted in the organism, provokes a response, the terminal step of which is usually, if not always, a reaction upon the external world" (Lotka 1925, 340). Second, his model uses the notion of information as controlling the behavior of an animal, and his simplistic model of representation (called "depiction" by Lotka) predates later work on cognitive representations. This notion allows talk of how the animal's behavior can be influenced by future events—represented thanks to the antennas mounted on the mechanical robot to detect the edge of a table—without appealing to the notion of the final cause (Cordeschi 2002, 128).

4. Computation and cybernetics: from Babbage and Turing to Ashby

Another crucial development in the nineteenth century was the blueprint for a general-purpose computational machine. Charles Babbage argued that his Difference Engine (see Figure 6.3) could serve as a simple model of the geological laws (Babbage 1837; Dolan 1998). This was one of the first uses of numerical computation for simulation purposes, but the implications of the Difference Engine do not end here. It was the predecessor of the Analytic Engine, the first general-purpose programmable machine to be designed, as Babbage himself and Lady Ada Lovelace observed. And a general-purpose mechanism was exactly what Descartes thought was impossible.

It was only Alan Turing (1937) who brought the consequences of the general-purpose computing machines to light. Having formalized the notion of a computing machine, called later a *Turing machine*, he defined a universal machine that was able to operate like any other Turing machine just by operating on the description of that machine on its input tape (for more, see Chapter 33). However, universal Turing machines can only compute effectively computable procedures, so there are limits to their capacities. Turing referred to the important work by Kurt Gödel (1933) who proved that in the first-order predicate logic with elementary arithmetic,

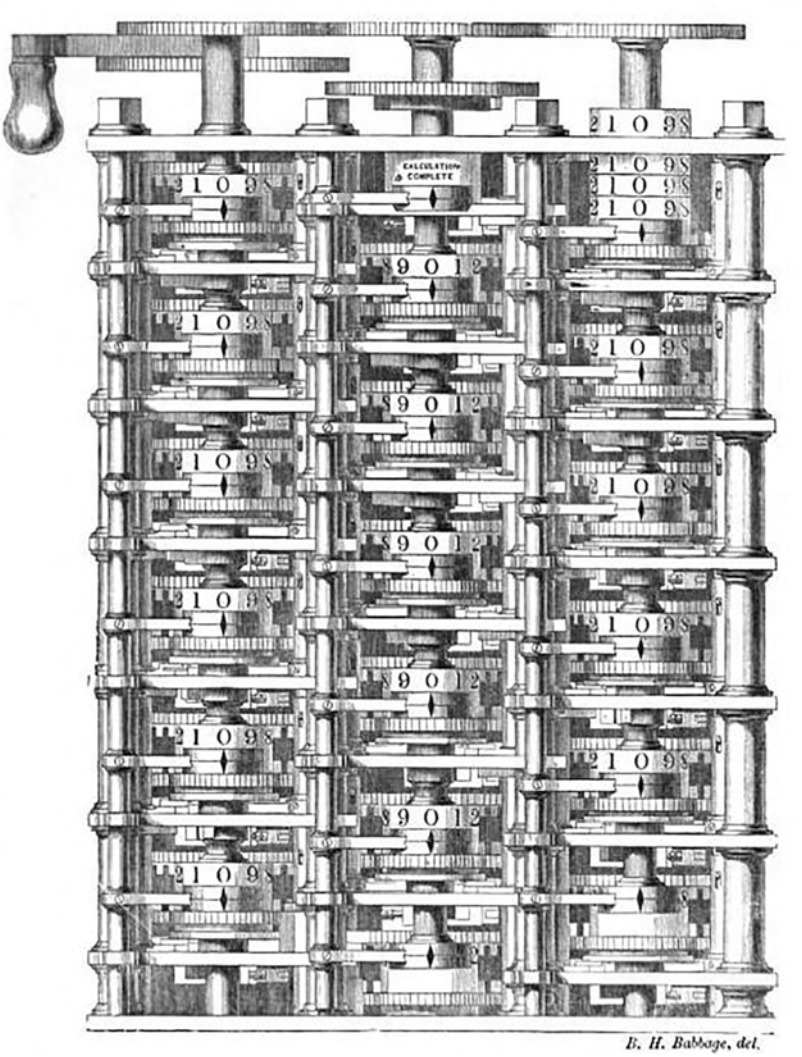

Figure 6.3 Difference Engine by Charles Babbage. Fragment of the original blueprint

there are always true statements that cannot be proved in that very logic. This later led to an antimechanistic argument that no machine could simulate a human mathematical mind (Lucas 1961; Penrose 1989). However, these arguments need to presuppose that the human mind is logically consistent, which also cannot be decided using an effective procedure (Putnam 1960). (The consistency presupposition is unjustified, and the proof used by Lucas leads to a contradiction (Krajewski 2007).)

Turing suggested that artificial computational machines can be thought to be generally intelligent as soon as they could converse with people without being detected as machines (Turing 1950). In this conversational test of intelligence—called later *the Turing test*—the criterion of thinking is what Descartes considered impossible.

Even more importantly, Turing was not alone when he thought machines could be programmed to display intelligence. It was a common assumption among British and American

defenders of *cybernetics*—a general theory of control and communication in the animal and the machine, as Norbert Wiener defined it in the eponymous title of his book (Wiener 1948).

Cybernetics made several important contributions to the development of mechanistic models of the mental (Pickering 2009; Johnston 2008). First, it offered a new conceptual framework for natural teleology, i.e. for describing the goals of physical systems without recourse to any non-physical factors (Rosenblueth, Wiener, and Bigelow 1943). The notion of feedback, fundamental to cybernetics, was used to elucidate the relation of the mechanism to its environment, showing how the environment can control the behavior of the mechanism. The *feedback* occurs when the output of the system is directed back to the input of the system, which effectively creates a loop. There are two kinds of feedback: negative and positive. The *negative feedback* occurs when some function of the output of the system is fed into its input to stabilize the output. It promotes stability and is frequently used for control purposes (Sluckin 1954). The *positive feedback*, on the other hand, occurs when a small disturbance in the output of the system increases the magnitude of the disturbance. It can lead to chaotic oscillations and exponential growth. Note that machines with feedback had been invented much earlier than the 1940s. There were other regulators or governors before (for example, the Watt governor used to control the steam engine), but there was no precise mathematical theory of control.

Second, cybernetics used the newly developed *theory of information* (Weaver and Shannon 1964). Claude Shannon defined a mathematical notion of information as uncertainty of the receiver of the information as to what information it will receive over the noisy channel from a sender (for introductions to the theory of information, see Floridi 2010; Burgin 2009). It has become possible to quantify the information sent. However, the open question was how to relate this theory to meaningful (semantic) information. Semantic information is *about* something, in contrast to purely structural, probabilistic, or syntactic information, which boils down to the existence of mere regularities. A number of theorists proposed their own approaches to semantic information (MacKay 1969; Bar-Hillel 1955; Dretske 1981).

Third, the focus on complexity of control allowed the hypothesis that the main function of the brain is to control—in a stable way—the overall behavior of the animal. One particularly influential notion was that of *homeostasis* or keeping the internal conditions of the system relatively constant and stable. W. Ashby built a model of a homeostatic machine—Homeostat—that was supposed to show how complex systems adapt to the environment (Ashby 1960). Even if the Homeostat does not seem particularly exciting (it has no effectors to move it), it shows what Ashby called "ultrastable behavior": it adjusts its internal organization to react appropriately to the outside disturbance. As such, Homeostat goes beyond simple reflex arc models and simple animal stimulus-driven behaviors.

Fourth, cybernetics promoted interdisciplinary research and asked integrative questions about mechanisms of control and communication in animals and machines. This promoted building robotic models of animals, and the use of cybernetic notions in biology and psychology. For example, W. Grey Walter built robots (Walter 1950a, 1950b, 1953)—called tortoises for their looks—as models of animal behavior: environmentally induced exploration (in a robot jokingly called *Machina speculatrix*); and conditional reflexes (in *Machina docilis*). These electromechanical tortoises were milestones in the history of biorobotics (the discipline that explains biological behavior with robots; compare Webb 2002) and bionics (the discipline that builds biologically inspired robots). The technological obstacles to building more complex robots could not, however, be easily surpassed, and robots gave way to computational modeling in information-processing psychology. Yet Grey Walter's research remained influential to defenders of classical cognitive research (Miłkowski 2016), and later, in the 1980s, inspired behavioral robotics (Brooks 1999).

5. Computational models and artificial intelligence

Cybernetics, information theory, and the advent of digital computers made it possible to consider the development of computer simulation of mental processes. While in the nineteenth and early twentieth century, Charles Sherrington, William James, and Alfred Lotka would talk of the telephone switchboard as the model of the brain, in the second half of the twentieth century the preferred mechanistic model of the mental was the programmable digital computer. There were four reasons given for the general analogy between computers and minds (Apter 1970). First, both are general-purpose machines. Second, both process information. Third, both can create *models* of reality. In other words, one of the basic functions of the brain is to build models of the environment; this was particularly stressed by Kenneth Craik (1967). Fourth, both computers and brains have a large number of similar elementary building blocks that perform simple operations, which, taken together, contribute to complex behavior.

In 1958, Herbert Simon and Allen Newell predicted that in ten years most psychological theories will take the form of "computer programs, or of qualitative statements about the characteristics of computer programs" (Simon and Newell 1958, 7–8). Their view was extreme, and many cognitive psychologists did not share their enthusiasm (Baars 1986; Neisser 1963). Nevertheless, psychologists used information-theory notions to design and analyze significant experimental results (Sperling 1960; Miller 1956; Broadbent 1958). Even in 1967, almost ten years after Simon and Newell made their initial prediction, psychological references to computer programs were highly metaphorical and lacked experimental evidence (Miller, Galanter, and Pribram 1967). In the 2000s, however, the prediction holds true: over 80 percent of articles in theoretical journals focus on computational models (Busemeyer and Diederich 2010).

What made the prediction true? One factor was the new interdisciplinary field formed at the end of the 1970s that came to be called *cognitive science* (Arbib et al. 1978). It may be a stretch to claim that it was a scientific revolution in the strict sense (Kuhn 1970), as some initially claimed (Gardner 1985): first, it does not seem incommensurable with previous psychological research (for example, Tolman's (1948) claim that rats use cognitive maps to navigate mazes is easily translatable into cognitive terminology); and second, it's deeply connected with multiple disciplines that already coalesced around cybernetics. Cognitive science made significant contributions to mechanistic models of the mental even if some would deny that all cognitive models are mechanistic.

In the 1950s, another important field was formed: Artificial Intelligence (AI). In AI, two basic approaches were dominant: connectionist machine learning, initially in the Perceptron machine (Rosenblatt 1958), and the rule-based approach, defended by Simon and Newell. The Perceptron was an electromechanical, artificial neural network (ANN) with extremely simplified neurons. It could learn by the so-called delta rule that allowed researchers to adjust the weights of connections between neurons. However, it was soon discovered that using the delta rule, Perceptrons could not learn elementary logical operations such as exclusive disjunction, or XOR (Minsky and Papert 1969). This has led ANN research to stall until the 1980s. In their place, rule-based approaches dominated, but they also failed to quickly deliver the promised results.

These two approaches have their counterparts in cognitive science. The rule-based approach, also called the *symbolic* approach because of its reliance on a particular conception of cognitive representation, was the mainstream approach until the early 1980s. For example, Allen Newell and Herbert Simon (1972) simulated problem-solving processes with rule-based systems, which were then evaluated by comparing the simulation with the behavioral data (such as verbal reports of subjects and eye-tracking data). Newell and Simon stressed that their methodology was ill-suited for motor and perceptual processing.

The connectionist approach has its roots in associationism, and it became much more sophisticated in the 1980s when new methods of ANN learning appeared. Notably, the backpropagation algorithm allowed the training of networks with multiple layers of hidden neurons, and the ANNs were used to simulate parallel cognitive processing. Significant progress has been made in transcending the limitations of associationism: the association relation is symmetrical while many cognitive classical structures are ordered (McClelland, Rumelhart, and PDP Research Group 1986; Rumelhart and McClelland 1986; for a general assessment, see Bechtel and Abrahamsen 2002). However, defenders of the symbolic approach claimed that ANNs either do not offer a full account of cognitive structures, with such features as compositionality or systematicity, or they are mere implementations of symbolic approaches (Fodor and Pylyshyn 1988).

All this criticism notwithstanding, ANN modeling has matured significantly. While the early ANNs were extremely simplified and idealized, the current models use much more biologically plausible models of neurons, sometimes with a large variety of neuron types. ANNs are also routinely described using control theory, which allows using dynamical systems theory to design their architecture (Eliasmith and Anderson 2003). They are also used in computational neuroscience, for example to build large brain models (Eliasmith 2013). The large models strive for biological plausibility (in the mechanistic terminology, they are how-plausibly models; see Craver 2007), and while they are heavily idealized, they are able to offer novel predictions. In other words, they are no longer just proofs of possibility that a mechanism implementing a mental capacity may be built, but idealized models of how actual brains may work.

Because of the heavy stress on computational models, general objections against cognitive science were raised. The first one is related to the notion of mental representation as used in cognitive research. In most explanatory models, it was assumed that it's sufficient to simulate the syntactic structure of representations for the semantic properties to appear; as John Haugeland quipped: "if you take care of syntax, the semantics will take care of itself" (Haugeland 1985, 106). John Searle raised an important objection against this formalist principle in his Chinese Room thought experiment (Searle 1980): one can easily imagine a person with a special set of instructions in English who could manipulate Chinese symbols and answer questions in Chinese without understanding it at all. Hence, understanding is not reducible to syntactic manipulation. While the discussion around this thought experiment is hardly conclusive (Preston and Bishop 2002), the problem was soon reformulated by Stevan Harnad as the "symbol grounding problem" (Harnad 1990). How can symbols in computational machines mean anything? Or, to spell this question out a little more generally, what mechanism is able to mean anything?

Formulated this way, Searle's question is yet another version of the mind–body problem as related to intentionality, or aboutness of mental processes or states. In the nineteenth century, Franz Brentano claimed that intentionality distinguishes the mental from the physical (Brentano 1900; Chrudzimski 1999). In this, he followed Descartes: no mechanism could entertain thoughts. However, with the advent of digital computers, a new argument was put forward: thought and content are causally relevant in the physical world just because syntactic entities are computed over. Searle argues that syntax is not sufficient for intentionality, and in this, most theorists of mental representation agree. The most influential position, teleosemantics, defended by Ruth Garett Millikan (1984) and Fred Dretske (1997), is that to represent is, roughly, to have the biological function of making information available to cognitive and motor processes. This position remains controversial (Ramsey 2007; Hutto and Myin 2013; Fodor and Pylyshyn 2015), especially for the followers of the cybernetic approach to cognition in terms of control. They claim that the appeal to (cognitive) representation in control mechanisms plays no particular explanatory role (Van Gelder 1995; Chemero 2009), and that computational explanations

may be replaced by dynamical explanations (essentially similar to the ones offered by Ashby in his account of Homeostat). However, most contemporary criticisms of computational cognitive science focus on two other properties of mental processes: qualities of experience and the ability to use common sense.

It was argued that no computational mechanism can produce the qualitative states of consciousness (called *qualia*) and that consciousness is not physical (Jackson 1986; Chalmers 1996). These arguments are usually put forward in the following form: no amount of physical knowledge (for example, about computational structures of the brain) is sufficient to know the qualitative consciousness. The two most frequent replies to such arguments are the following: first, it is claimed that qualitative knowledge of conscious states is not knowledge-that but knowledge-how (Ryle 2009), which is in principle not available intersubjectively (Churchland 1985); and second, it is argued that it's just the lack of imagination of philosophers trying to understand the epistemic situation of a person with all the relevant physical knowledge that creates the illusion that there is something to qualia that is beyond all physical description (Dennett 2005).

The last important argument against computational cognitive science is that it cannot explain how people smoothly cope with their everyday tasks, or to understand their common sense (Dreyfus 1979). Indeed, it is still a Holy Grail of artificial intelligence to build a machine that can make commonsensical inferences and act accordingly. The defenders of AI claim that the actual scope of common-sense knowledge is vaster than critics of AI imagine.

6. Conclusion

The history of mechanistic models of the mental reflects theorizing about mechanisms as such. First, mechanisms were considered to be entities in space, whose changes could be described geometrically, and whose interactions could be only local. Then, they were considered causal and temporal as well. La Mettrie subsequently stressed that mechanisms are organized systems. At the same time, the understanding of the mental in terms of mechanisms of information processing progressed. Descartes proposed a general outline of the nervous system in terms of a reflex arc but considered important mental faculties to be exempt from the purview of mechanism. He supposed that there could be no mechanism for thinking and drawing inferences. Soon, however, Leibniz sketched a proposal of a universal, logical, and computational calculus of thought. As soon as computing mechanisms were built, it was suggested that they could be used to build not only rough approximations of the mind, but also precise causal models, including also cyclical interactions with the environment. Subsequently, the gap between mathematical, computational simulations of the mental and the neurobiologically plausible models began to close.

The role of mechanistic models has also changed. Here, it is useful to distinguish mechanistic simulations from theoretical descriptions of mechanisms. Until the second half of the twentieth century, mechanistic simulations, usually owing to limits of technology, were mere proofs of possibility that the mind might be, in spite of appearances, mechanical. Some modelers were overly enthusiastic about the capacities of their artifacts; for example, Walter would claim that his tortoises were capable of self-recognition, while they simply responded to light sources, and just because they had lamps mounted on them, they would react to their own image in the mirror. Theoretical descriptions of mechanisms were not bounded by technological limitations, and they could offer some insight into how the mental processes actually occur. Hence, already the model of the reflex arc was not just a mere how-possibly model that would prove that the mechanical mind is possible (and Descartes would even retort that the mechanical mind is a contradiction in terms). It was a how-plausibly model of the nervous system (not the mind) whose

role was to explain at least a large number of muscle movements. Only in the second half of the twentieth century have mechanistic simulations been built for similar explanatory purposes. Yet technological and theoretical limitations have not all been overcome, and they are still far from complete models of how actual mechanisms work. But the underlying assumption of modelers in cognitive neuroscience today is that such models not only can but also should be built (Piccinini and Craver 2011; Miłkowski 2013; Boone and Piccinini 2016).

Different technological artifacts have been treated as models of the mental, from a wax tablet, through clocks and the automated telephone switchboard, to a digital computer. However, as Margaret Boden stresses, the computer is not just another metaphor of the mind; it might also be the last metaphor because it may offer a precise causal model of the mind (Boden 2008). Most mechanistic models of the mental are no longer controversial. Even self-proclaimed dualists rarely believe that it is impossible to understand intentionality mechanistically. The only remaining bastion of anti-mechanists is qualitative consciousness, although in recent years, some progress has been made in terms of measuring consciousness (Seth et al. 2008) and building more plausible causal models of its processes (Boly et al. 2013). We still do not know how to model consciousness mechanistically, but it is no longer obvious that we will never know.

References

Apter, Michael. 1970. *The Computer Simulation of Behaviour*. London: Hutchinson.

Arbib, Michael A., Carl Lee Baker, Joan Bresnan, Roy G. D'Andrade, Ronald Kaplan, Samuel Jay Keyser, Donald A. Norman, et al. 1978. *Cognitive Science, 1978*: *Report of The State of the Art Committee to The Advisors of The Alfred P. Sloan Foundation*. New York: The Alfred P. Sloan Foundation. http://csjarchive.cogsci.rpi.edu/misc/CognitiveScience1978_OCR.pdf.

Ashby, W Ross. 1960. *Design for a Brain: The Origin of Adaptive Behavior*. New York: Wiley.

Baars, Bernard J. 1986. *The Cognitive Revolution in Psychology*. New York: Guilford Press.

Babbage, Charles. 1837. *The Ninth Bridgewater Treatise. A Fragment*. London: J. Murray.

Bar-Hillel, Yehoshua. 1955. "Information and Content: A Semantic Analysis." *Synthese* 9 (1): 299–305. doi:10.1007/BF00567416.

Bechtel, William and Adele Abrahamsen. 2002. *Connectionism and the Mind*. Oxford: Blackwell.

Bilder, Robert and F. Frank LeFever, eds. 1998. *Neuroscience of the Mind on the Centennial of Freud's Project for a Scientific Psychology*. New York: New York Academy of Sciences.

Boden, Margaret A. 2008. "Information, Computation, and Cognitive Science." In *Philosophy of Information: Volume 8*, edited by Pieter Adriaans and Johan Van Benthem, 8: 741–61. Elsevier B.V. doi:10.1016/B978-0-444-51726-5.50023-6.

Boly, Melanie, Anil K. Seth, Melanie Wilke, Paul Ingmundson, Bernard J. Baars, Steven Laureys, David Edelman, and Naotsugu Tsuchiya. 2013. "Consciousness in Humans and Non-Human Animals: Recent Advances and Future Directions." *Frontiers in Psychology* 4. Frontiers. doi:10.3389/fpsyg.2013.00625.

Boone, Worth and Gualtiero Piccinini. 2016. "The Cognitive Neuroscience Revolution." *Synthese*, June 193 (3): 1509–1534. Springer Netherlands. doi:10.1007/s11229-015-0783-4.

Brentano, Franz. 1900. *Psychologie Vom Empirischen Standpunkt*. Hamburg: F. Meiner.

Broadbent, D. E. 1958. *Perception and Communication*. Oxford: Pergamon Press.

Brooks, Rodney A. 1999. *Cambrian Intelligence: The Early History of the New AI*. Cambridge, MA: The MIT Press.

Burgin, Mark. 2009. *Theory of Information: Fundamentality, Diversity and Unification*. Singapore: World Scientific Publishing Co.

Burke, Robert E. 2007. "Sir Charles Sherrington's the Integrative Action of the Nervous System: A Centenary Appreciation." *Brain: A Journal of Neurology* 130 (Pt 4): 887–94. doi:10.1093/brain/awm022.

Busemeyer, Jerome R. and Adele Diederich. 2010. *Cognitive Modeling*. Los Angeles: Sage.

Canguilhem, Georges. 1955. *La Formation Du Concept de Réflexe Aux XVIIe et XVIIIe Siècles*. Paris: Presses universitaires de France.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chemero, Anthony. 2009. *Radical Embodied Cognitive Science*. Cambridge, MA: The MIT Press.

Chrudzimski, Arkadiusz. 1999. "Die Theorie Der Intentionalität Bei Franz Brentano." *Grazer Philosophische Studien* 57 (July): 45–66. doi:10.5840/gps1999574.

Churchland, Paul M. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States." *The Journal of Philosophy* 82 (1): 8–28.

Cordeschi, Roberto. 2002. *The Discovery of the Artificial*. Studies in Cognitive Systems. Dordrecht: Springer Netherlands. doi:10.1007/978-94-015-9870-5.

Craik, Kenneth. 1967. *The Nature of Explanation*. Cambridge: Cambridge University Press.

Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Dennett, Daniel C. 2005. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: The MIT Press.

Descartes, René. 1985. *The Philosophical Writings of Descartes*. Edited by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Vol. 1. Cambridge: Cambridge University Press.

Dolan, Brian P. 1998. "Representing Novelty: Charles Babbage, Charles Lyell, and Experiments in Early Victorian Geology." *History of Science* 36: xxxvi.

Dretske, Fred I. 1981. *Knowledge and the Flow of Information*. 2nd ed. Cambridge, MA: The MIT Press.

———. 1997. *Naturalizing the Mind*. Cambridge, MA: The MIT Press.

Dreyfus, Hubert. 1979. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: The MIT Press.

Eliasmith, Chris. 2013. *How to Build the Brain: A Neural Architecture for Biological Cognition*. New York: Oxford University Press.

Eliasmith, Chris and Charles H. Anderson. 2003. *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. Cambridge, MA: The MIT Press.

Floridi, Luciano. 2010. *Information: A Very Short Introduction*. Oxford: Oxford University Press.

Fodor, Jerry A. and Zenon W. Pylyshyn. 1988. "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition* 28 (1–2): 3–71.

———. 2015. *Minds without Meanings: An Essay on the Content of Concepts*. Cambridge, MA: The MIT Press.

Gardner, Howard. 1985. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Gödel, Kurt. 1933. "Zum Entscheidungsproblem Des Logischen Funktionenkalküls." *Monatshefte Für Mathematik Und Physik* 40 (1): 433–43. doi:10.1007/BF01708881.

Harnad, Stevan. 1990. "The Symbol Grounding Problem." *Physica D* 42: 335–46.

Haugeland, John. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, MA: The MIT Press.

Hobbes, Thomas. 1839. *The English Works of Thomas Hobbes of Malmesbury*. London: J. Bohn.

Hutto, Daniel D. and Erik Myin. 2013. *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: The MIT Press.

Jackson, Frank. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83 (5): 291–5. doi:10.2307/2026143.

Johnston, John. 2008. *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*. Cambridge, MA: The MIT Press.

Krajewski, Stanisław. 2007. "On Gödel's Theorem and Mechanism: Inconsistency or Unsoundness Is Unavoidable in Any Attempt to 'Out-Gödel' the Mechanist." *Fundamenta Informaticae* 81 (1): 173–81.

Krueger, Robert C. and Clark L. Hull. 1931. "An Electro-Chemical Parallel to the Conditioned Reflex." *Journal of General Psychology* 5: 262–9.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

La Mettrie, Julien. 1912. *Man a Machine*. Translated by Gertrude Carman Bussey and Mary Whiton Calkins. Chicago: The Open Court Pub. Co.

Leibniz, Gottfried. 1666. *Dissertatio de Arte Combinatoria, in qua Ex Arithmeticae Fundamentis Complicationum Ac Transpositionum Doctrina Novis Praeceptis Extruitur, & Usus Ambarum per Universum Scientiarum Orbem Ostenditur*. Lipsiae: apud Joh. Simon Fickium et Joh. Polycarp. Seuboldum Literis Spörelianis.

Lotka, Alfred. 1925. *Elements of Physical Biology*. Baltimore: Williams & Wilkins Company.

Lucas, J. R. 1961. "Minds, Machines and Gödel." *Philosophy* 9 (3): 219–27.

MacKay, Donald MacCrimmon. 1969. *Information, Mechanism and Meaning*. Cambridge, MA: The MIT Press.

McClelland, James L., David E. Rumelhart, and PDP Research Group, eds. 1986. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Volume 2: Psychological and Biological Models*. Cambridge, MA: The MIT Press.

Miller, George A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (2): 81–97. doi:10.1037/h0043158.
Miller, George A., Eugene Galanter, and Karl H. Pribram. 1967. *Plans and the Structure of Behavior*. New York: Holt.
Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: The MIT Press.
Miłkowski, Marcin. 2013. *Explaining the Computational Mind*. Cambridge, MA: The MIT Press.
——. 2016. "Models of Environment." In *Minds, Models and Milieux: Commemorating the Centennial of the Birth of Herbert Simon*, edited by Roger Frantz and Leslie Marsh, 227–38. New York: Palgrave Macmillan. doi:10.1057/9781137442505_13.
Minsky, Marvin and Seymour Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: The MIT Press.
Neisser, U. 1963. "The Imitation of Man by Machine: The View That Machines Will Think as Man Does Reveals Misunderstanding of the Nature of Human Thought." *Science* 139 (3551): 193–7. doi:10.1126/science.139.3551.193.
Newell, Allen and Herbert A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
Penrose, Roger. 1989. *The Emperor's New Mind*. London: Oxford University Press.
Piccinini, Gualtiero and Carl F. Craver. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183 (3): 283–311. doi:10.1007/s11229-011-9898-4.
Pickering, Andrew. 2009. *The Cybernetic Brain: Sketches of Another Future*. Chicago: University of Chicago Press.
Preston, John and Mark Bishop. 2002. *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford: Clarendon Press.
Putnam, Hilary. 1960. "Minds and Machines." In *Dimensions of Mind*, edited by Sidney Hook, 148–79. New York: New York University Press.
Ramsey, William M. 2007. *Representation Reconsidered*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511597954.
Rosenblatt, F. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review* 65 (6): 386–408. doi:10.1037/h0042519.
Rosenblueth, Arturo, Norbert Wiener, and Julian Bigelow. 1943. "Behavior, Purpose and Teleology." *Philosophy of Science* 10 (1): 18. doi:10.1086/286788.
Rumelhart, D. E. and James L. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1. Foundations*. Cambridge, MA: MIT Press.
Russell, S. Bent. 1913. "A Practical Device to Stimulate the Working of Nervous Discharges." *Journal of Animal Behavior* 3 (1): 15–35. doi:10.1037/h0070584.
Ryle, Gilbert. 2009. *The Concept of Mind*. New York: Routledge.
Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 1–19. doi:10.1017/S0140525X00005756.
Seth, Anil K., Zoltán Dienes, Axel Cleeremans, Morten Overgaard, and Luiz Pessoa. 2008. "Measuring Consciousness: Relating Behavioural and Neurophysiological Approaches." *Trends in Cognitive Sciences* 12 (8): 314–21. doi:10.1016/j.tics.2008.04.008.
Simon, Herbert A. and Allen Newell. 1958. "Heuristic Problem Solving: The Next Advance in Operations Research." *Operations Research* 6 (1): 1–10.
Sluckin, Wladyslaw. 1954. *Minds and Machines*. Harmondsworth: Penguin.
Sperling, George. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74 (11): 1–29. doi:10.1037/h0093759.
Thorndike, Edward L. 1911. *Animal Intelligence*. London: The Macmillan Company.
Tolman, Edward Chace. 1939. "Prediction of Vicarious Trial and Error by Means of the Schematic Sowbug." *Psychological Review* 46 (4): 318–36. doi:10.1037/h0057054.
——. 1948. "Cognitive Maps in Rats and Men." *Psychological Review* 55 (4): 189–208.
Turing, Alan. 1937. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* s2-42 (1): 230–65. doi:10.1112/plms/s2-42.1.230.
——. 1950. "Computing Machinery and Intelligence." *Mind* LIX (236): 433–60. doi:10.1093/mind/LIX.236.433.
Van Gelder, T. 1995. "What Might Cognition Be, If Not Computation?" *The Journal of Philosophy* 92 (7): 345–81.
Walter, W. Grey. 1950a. "An Imitation of Life." *Scientific American*. doi:10.1038/scientificamerican0550-42.
——. 1950b. "An Electro-Mechanical Animal." *Dialectica* 4 (3): 206–13. doi:10.1111/j.1746-8361.1950.tb01020.x.

——. 1953. *The Living Brain*. New York: Norton.

Weaver, W. and C. E. Shannon. 1964. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Webb, Barbara. 2002. "Can Robots Make Good Models of Biological Behaviour?" *Behavioral and Brain Sciences* 24 (6): 1033–94. doi:10.1017/S0140525X01000127.

Wheeler, Michael. 2008. "God's Machines: Descartes on the Mechanization of Mind." In *The Mechanical Mind in History*, edited by Phil Husbands, Owen Holland, and Michael Wheeler, 307–30. Cambridge, MA: The MIT Press.

Wiener, Norbert. 1948. *Cybernetics, or Control and Communication in the Animal and the Machine*. New York/Paris: J. Wiley/Hermann.

Willis, Thomas. 1694. *Opera Omnia, Nitidius Quam Unquam Hactenus Edita, Plurimum Emendata*. Coloniae: Sumptibus Gasparis Storti.