

COPYRIGHT NOTICE:

Jean-Pierre Dupuy: The Mechanization of the Mind

is published by Princeton University Press and copyrighted, © 2000, by Princeton University Press. All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher, except for reading and browsing via the World Wide Web. Users are not permitted to mount this file on any network servers.

For COURSE PACK and other PERMISSIONS, refer to entry on previous page. For more information, send e-mail to permissions@pupress.princeton.edu

The Self-Mechanized Mind

FROM 1946 TO 1953 ten conferences—the first nine held at the Beekman Hotel at 575 Park Avenue in New York, the last at the Nassau Inn in Princeton, New Jersey—brought together at regular intervals some of the greatest minds of the twentieth century. Sponsored by the Josiah Macy, Jr. Foundation, these meetings have since come to be known as the Macy Conferences. The mathematicians, logicians, engineers, physiologists, neurophysiologists, psychologists, anthropologists, and economists who took part set themselves the task of constructing a general science of how the human mind works. What brought them together, what they discussed, and what came of a collaboration unique in the history of ideas—these things form the subject of the present volume.

Every group of this kind adopts a code name as a way of affirming its identity. In the case of the Macy Conferences it was “cybernetics.” Today this name has fallen out of fashion, to say the least. Since 1954 the project undertaken by the Cybernetics Group¹ has been carried on under a series of different names, ultimately coming to be known as “cognitive science.” Why cognitive science today is ashamed of its cybernetic heritage is one of the chief questions I wish to address.

The Cybernetic Credo

The Cybernetics Group drew exceptional energy and passion from two convictions that were shared by most of its members and that were so novel at the time that the simple fact of defending them made one part of an elitist avant-garde, worshipped by some and demonized by others. These two convictions were based on logical and scientific discoveries that had been made in the immediately preceding decades, the consequences of which the members of the Cybernetics Group intended to exploit to the fullest. In very general terms, which will need subsequently to be made more precise, they held that:

1. Thinking is a form of computation. The computation involved is not the mental operation of a human being who manipulates symbols in applying

rules, such as those of addition or multiplication; instead it is what a particular class of machines do—machines technically referred to as “algorithms.” By virtue of this, thinking comes within the domain of the mechanical.

2. Physical laws can explain why and how nature—in certain of its manifestations, not restricted exclusively to the human world—appears to us to contain meaning, finality, directionality, and intentionality.

Inspired by these two articles of faith with a fervor and an enthusiasm rarely matched in the history of science, the founders of the cybernetic movement believed they were in a position to achieve very great things. They thought it possible to construct a scientific, physicalist theory of the mind, and thereby resolve the ancient philosophical problem of mind and matter. They thought themselves capable of reconciling the world of meaning with the world of physical laws. Thanks to them, the mind would at last find its rightful place in nature.

They were neither the first nor the last to conceive of such an ambition. In the past it had generally been philosophers who debated such questions, fraught with metaphysical assumptions and implications. Materialists feared that the slightest hint of dualism would let back in religion, which they abhorred; dualists, for their part, saw in materialism a threat to man’s free will. But now scientists and engineers ventured to address what formerly had been philosophical problems. For them, to devise a theory meant to build a model, which was to be constructed not only on paper but in physical form as well. They were persuaded that one truly understands only that which one has made, or is capable of making. Their research program would be realized, they thought, only when—like God, who was supposed to have created the universe—they had succeeded in making a brain that exhibited all the properties associated with the mind. Indeed, one of the most influential works of the cybernetics movement bore the title *Design for a Brain*.²

In this book I retrace the exceedingly complex intellectual history of this movement and defend the thesis that contemporary cognitive science is fully rooted in cybernetics. This is not to say that the ideas underlying cybernetics do not differ from those of the rival paradigms that contend within cognitive science today. Ideas have changed to such an extent during the past half-century, in fact, that I feel the need to warn the reader at the outset against a number of serious misunderstandings on the part of cognitive scientists that are the result of interpreting cybernetics in the light of present-day conceptions. It will be well, then, to go back to the two basic convictions mentioned earlier and to examine how they differ from current beliefs.

The cyberneticians’ first thesis—that to think is to compute as a certain class of machines do—amounted to analyzing and describing what it is to

think, not, as it is commonly supposed, to deciding whether it is possible to conceive of machines that think. The question “Can a machine think?” did not come to the forefront until later, at the beginning of the 1950s, as the research program known as artificial intelligence began gradually to establish itself within computer science. To this question cybernetics obviously could only respond in the affirmative, since it had already defined the activity of thinking as the property of a certain class of machines. It is important to see, however, that cybernetics represented not the anthropomorphization of the machine but rather the mechanization of the human. This is only one of many received ideas about cybernetics that needs to be stood on its head.

When the question “Can a machine think?” is posed today, one thinks above all of computers. Cognitivism—as the tendency that has long been dominant within cognitive science is known—is often thought of as dogmatically relying upon the computer metaphor: to think is to manipulate physical symbols by following rules after the fashion of a computer program. This definition is commonly—and altogether mistakenly—said to be due to cybernetics. The error has to do, first of all, with a matter of historical fact: when the cybernetics movement came into being, the computer did not yet exist. As we shall see, the computer was conceived by John von Neumann as a direct result of cybernetic ideas; it did not form the technological background against which these ideas developed. Here again the reversal of perspective that needs to be brought about is total. In this case, at least, the old “idealist” thesis—so denigrated by sociologists of science, who see themselves as resolute materialists—turns out to have been correct: it is not the physical world that determines the evolution of ideas, but rather ideas that generate scientific and technological development.

But the error committed by those who hold cybernetics responsible for identifying thinking with the functioning of a computer is above all a philosophical error, the nature of which it is important to grasp. The computations carried out by a computer are of a very special kind in that they involve symbols, which is to say representations. On the cognitivist view, symbols are objects that have three aspects: physical, syntactic, semantic. It is on the strength of these symbols that cognitivism claims to be able to span the gap that separates the physical world from the world of meaning. Computation may therefore be described as the central pier of the cognitivist bridge. Cognitivism also assumes that purely formal computations carried out at the syntactic level are materially embodied in the causal processes that occur within the computer insofar as it is a physical object, and are interpreted at the semantic level on the basis of elementary meanings assigned to the symbols. This solution to the problem posed by the presence of meaning in a world of physical facts that are linked by causal

laws has been sharply attacked. As John Searle, one of the fiercest and most influential critics of cognitivism and artificial intelligence, has argued, “Syntax by itself is neither constitutive of nor sufficient for semantics,”³ which is to say that the execution of a computer program cannot in principle enable the machine to understand what it does, to be conscious of what it does, or to give meaning to the world in which it functions. The cognitivists have replied by conceding that a computer program, being an abstract and purely syntactic object, naturally cannot claim to be equipped with a mind, much less claim to *be* a mind. They concede also that mind can arise in the physical world only from a causal dynamics. But cognitivism asserts that if a mind arises as a result of implementing a certain program in the physical world, then any implementation of the same program in a different hardware, no matter what it may be, would produce a mind endowed with the same properties. In other words, what is essential for the emergence of mind is not the concrete causal organization of this or that material system possessing a mind; what is essential is its *abstract* causal organization, which remains invariant when one passes from one material system to another.⁴

The fundamental concepts that allow cognitivism to advance this type of argument today are due to cybernetics. As it happens, however, cybernetics conceived and articulated these concepts in a very different manner that made no reference whatever to a computer—an object that had yet to be invented in the form in which we now know it. The three levels postulated by cognitive science—computation, causal physical laws, and meaning—were already developed in cybernetic thinking; but, in passing from cybernetics to cognitivism, both the character and the order of these levels came to be altered. How and why this transformation came about is another one of the questions that I will address in the course of this book.

First of all, computation as cybernetics conceived it is not symbolic computation; that is, computation involving representations. It is purely “mechanical,” devoid of meaning. The objects on which cybernetic computation bears have no symbolic value, being carried out by a network of idealized neurons in which each neuron is an elementary calculator that computes zeroes and ones as a function of the signals it receives from the neurons with which it is in communication. This type of neuronal network, which mimics the anatomical structure and functional organization of the brain, is one of cybernetics’ very greatest conceptual inventions. Under the name of “connectionism” it has since come to constitute within contemporary cognitive science, and in particular within the field of artificial intelligence, a fully fledged paradigm capable of competing with cognitivism and what is now called classical artificial intelligence. A history of cognitive science that omits mention of its cybernetic origins, as is often done, gives the impression that connectionism is a new paradigm, devised

relatively recently in order to rescue cognitive science from the impasses into which cognitivism had led it. This again is a glaring error. Once cybernetics is reintegrated into the history of cognitive science, as it must be, it becomes clear that computation was first introduced into the construction of a materialist and physicalist science of the mind not as symbolic computation involving representations, but instead as a sort of blind computation having no meaning whatever, either with respect to its objects or to its aims.

And if, in the cybernetic conception, meaning and mind happen to be associated with matter, it is because they *arise from* it. Someone who subscribes to Searle's critique of cognitivism and who holds that symbolic computation is incapable of giving rise to meaning, even though it concerns objects that already have a semantic value, may feel all the more strongly tempted to conclude that a type of computation that is devoid of any meaning whatsoever has still less chance of conjuring up meaning. But this is just the point: the cyberneticians did not derive meaning from computation; they derived it from causal physical laws. This brings us to the second of the basic convictions of cybernetics.

As we enter the twenty-first century, there is nothing in the least odd about the idea of a physics of meaning.⁵ An impressive series of scientific and mathematical discoveries made during the second half of the twentieth century has completely changed the way in which we conceive of dynamics, the branch of mechanics (formerly described as "rational") that concerns the path of development or trajectory of a material system subject to purely causal physical laws. It is well known today that complex systems, made up of many elements interacting in nonlinear ways, possess remarkable properties—so-called *emergent* properties—that justify their description in terms that one should have thought had been forever banished from science in the wake of the Galilean-Newtonian revolution. Thus it is said of these systems that they are endowed with "autonomy," that they are "self-organizing," that their paths "tend" toward "attractors," that they have "intentionality" and "directionality"—*as if* their paths were guided by an end that gives meaning and direction to them even though it has not yet been reached; as if, to borrow Aristotelian categories, purely efficient causes were capable of producing effects that mimic the effects of a final cause.

The many physico-mathematical concepts and theories that have contributed to this upheaval fit together with each other in extremely complicated ways. One thinks of "catastrophes," attractors and bifurcations of nonlinear dynamical systems, critical phenomena and symmetry breakings, self-organization and critical self-organizing states, nonlinear thermodynamics and dissipative structures, the physics of disordered systems, deterministic chaos, and so on. The models of this new physics make it

possible to understand the mechanisms of morphogenesis, which is to say the emergence of qualitative structures at a macroscopic level that organize themselves around the singularities—or qualitative discontinuities—of underlying processes at the microscopic level. By studying and classifying these singularities, which structure the way in which physical phenomena appear to us, it may be possible to construct a theory of meaning. As the editors of a recent collection of articles on this topic put it, meaning “is perfectly susceptible to a physicalist approach provided that we rely upon the qualitative macrophysics of complex systems and no longer upon the microphysics of elementary systems.”⁶ It is very telling in this respect that the physics to which cognitivism refers when it undertakes to “naturalize” and “physicalize” the mind remains precisely a microphysics of elementary systems. Indeed, it would not be unfair to say that by postulating an ultimate microlevel of reality whose elements are subject to fundamental laws, the cognitivists’ physics is essentially a physics of philosophers, evidence for which is nowhere to be found today in actual physics laboratories.

What has cybernetics got to do with all of this? To be sure, the physical and mathematical theories just mentioned did not yet exist in the early days of cybernetics, or were then only in an embryonic stage. But the cyberneticians, most of them outstanding specialists in physics and mathematics, were armed with a battery of concepts that included not only the notion—already classical at the time—of an attractor of a dynamical system, but also more revolutionary notions that they invented or at least considerably developed, such as feedback, circular causality, systems, and complexity. Above all they disposed of an incomparable theoretical instrument: the neural network. These things were quite enough to fortify the cyberneticians in what I described at the outset as their second fundamental conviction. Eloquent testimony to this is the fact that before adopting the name “cybernetics,” the movement described its mission as the elaboration of a “theory of teleological mechanisms.” Indeed the title the cyberneticians gave to their first meetings was just this: “Teleological Mechanisms”—an expression that is continually encountered in their writings and speeches, whether these were intended for publication in technical journals, for presentation at scientific conferences, or to sway financial backers and the general public. It is almost impossible today to imagine how very scandalous this formula seemed at the time. It appeared to be the ultimate oxymoron, a pure contradiction in terms, for it seemed to conflate two types of explanation: mechanical explanation by causes, the only kind admitted by science, and explanation by ends (*telos*), utterly forbidden by science. Naturally the cyberneticians did not think themselves guilty of any such confusion; naturally, too, they yielded completely to scientific

discipline in recognizing only causal explanations. What the expression “teleological mechanisms” was meant to signify was the capacity of certain complex physical systems, through their behavior, to mimic—to *simulate*—the manifestations of what in everyday language, unpurified by scientific rigor, we call purposes and ends, even intention and finality. The cyberneticians believed that behind these manifestations there lay only a causal organization of a certain type, which it was their business to identify. In other words, no matter the hardware in which this type of causal organization is implemented—and, in particular, no matter whether the hardware is part of a natural physical system or an artificial physical system—this causal organization will give off the same effects of purpose and intentionality.

John Searle, in his critique of cognitivism, has asserted that its principal error consists in confusing simulation and duplication. To quote one of his favorite examples, it would be absurd—as everyone will readily agree—to try to digest a pizza by running a computer program that simulates the biochemical processes that occur in the stomach of someone who actually digests a pizza. How is it then, Searle asks, that cognitivists do not see that it would be just as absurd to claim to be able to duplicate the neurobiological functioning of the mind by running a computer program that simulates, or models, this functioning? Fair enough. But how much force would this argument have if the process in question, rather than being a physical process such as digesting a pizza, were already itself a simulation? Consider the attempt by certain theorists (notably French deconstructionists) to demystify the concept of money. Noting that in the form of fiat money—paper currency, for example—it is a pure sign, lacking intrinsic value, inconvertible into gold or other mineral treasure, they conclude that money by its very nature is counterfeit. The fact that money nonetheless remains the basis of commercial exchange is due, they argue, solely to the existence of a potentially infinite chain of shared gullibilities: if money (truly) possesses a positive value, this is only because everyone (falsely) believes that it possesses a positive value. Let us suppose that this theory is correct: it must then be concluded that there is no essential difference between a dollar bill printed by the Federal Reserve and a *simulated* dollar bill—a counterfeit dollar. This counterfeit dollar will be used in commercial exchange in the same way as the dollar that has officially been authorized as legal tender, so long as it is believed to have the same value; that is, so long as no one suspects that it is counterfeit. For the cyberneticians, meaning is by its very nature counterfeit: its essence is confused with its appearance. To simulate this essence, for example by means of a model, is to remain true to it, since simulation amounts actually to duplicating it. This argument is one that the cyberneticians, for their part, could have

made in response to a critique of the sort that Searle brings against cognitivism; it is not, however, an argument that cognitivists can use to defend themselves against Searle's attacks.

Must it therefore be said that the cyberneticians reduced meaning and finality to a pure illusion produced by certain forms of causal organization? One might with equal justification hold that, to the contrary, they rescued phenomenality—appearance—by uncovering the mechanisms (the algorithms) that generate it. The cyberneticians themselves were divided between these two interpretations, depending on their sensibility. The most radical and uncompromising among them wholeheartedly embraced the project of demystifying appearance; others showed greater subtlety, implicitly adopting the strategy developed by Kant in the second part of his third Critique, the *Kritik der Urteilkraft*, entitled “Critique of Teleological Judgment.” In a sense, the typically cybernetic expression “teleological mechanisms” constitutes a striking condensation of this strategy. Only explanations that ultimately appeal to causal mechanisms are considered adequate. Nonetheless, faced with the most surprising manifestations of complexity in nature (life for Kant, the mind for the cyberneticians), recourse to another “maxim of judgment”—teleological judgment—becomes inevitable. Concepts such as “internal finality” are indispensable, and perfectly legitimate, so long as one keeps in mind that they have only heuristic and descriptive relevance. Teleological judgment consists in treating them *as though*—the Kantian *als ob*—they have objective value. The role played by simulation in the history of cognitive science since the earliest days of cybernetics is in part a reflection of this doctrine of make-believe.

Until now I have spoken of the cyberneticians as though they were a homogenous group, while suggesting a moment ago that there were clear differences of temperament and viewpoint within the movement. These differences were, in fact, sometimes quite pronounced. The debates, controversies, and, indeed, conflicts that resulted from them were what gave this episode in the history of ideas its exceptional richness. Whatever unity cybernetics may have enjoyed was a complicated one. The work of the Cybernetics Group, which constitutes the main object of this book, produced a very considerable progeny. In particular, it is important to note that it gave birth to a second cybernetics, very different in style than the first. This offshoot, which called itself “second-order cybernetics,” was founded by Heinz von Foerster, who in 1949 became the secretary of the Macy Conferences. From 1958 to 1976 its home was the Biological Computer Laboratory, established and directed by von Foerster on the campus of the University of Illinois at Urbana-Champaign. The place of the second cybernetics in the history of cognitive science is modest by contrast with the importance of the concepts it developed and with the influence that

these concepts were to exercise upon a great number of researchers (the present author included) in a wide variety of fields. One of its chief topics of research, self-organization in complex systems, led to fascinating breakthroughs in the direction of giving a physical interpretation to meaning. Its great misfortune was to have been overshadowed by artificial intelligence and cognitivism, which experienced a boom during these same years. Because von Foerster and his collaborators had the audacity—or perhaps only because they were foolish enough—to adopt in their turn the label “cybernetics,” which in the meantime had acquired a poor reputation, the leaders of these rival movements, now in the process of asserting their primacy, dismissed them as amateurs and nuisances. But the history of cognitive science is by no means finished. As a result of the success currently enjoyed by connectionism, one begins to see the first timid signs of renewed interest on the part of some researchers in the ideas of the second cybernetics. Although these ideas are not the main interest of the present book, I will refer to them frequently in trying to explain why the first cybernetics, confronted with the theories of self-organization and complexity that were to be dear to its successor, turned its back on them, and indeed sometimes—a cruel irony!—actually combatted them. In retrospect this appears to show an astonishing lack of lucidity on the part of the original cyberneticians. What is more, it suggests that the origins of cognitive science lie in a failure.

Cybernetics and Cognitivism

We are now in a position to draw up a preliminary list of those things that both united cognitivism with its cybernetic parent and led to the break between them. Three levels of analysis are present in each case: computation, physical causality, and meaning. But hidden behind the nominal identity of these terms lie great differences in interpretation. The physics of the cognitivists is a fictional physics, a philosopher’s physics; the physics of the cyberneticians is a true physics, a physicist’s physics. The computation of the cyberneticians is a pure computation, independent of any reference to meaning; the computation of the cognitivists is a symbolic, conceptual computation, closely linked to meaning. Finally, and perhaps most important, the meaning in which cognitivists are interested is rooted in the properties of beings endowed with intentionality, and possibly consciousness as well: human beings and societies, organized biological forms, living and complex artificial systems. The meaning with which the cyberneticians were concerned was much more universal and abstract—“structuralist,” one is tempted to say—in that it was consubstantial with all sufficiently complex forms of organization in nature, not excluding

inanimate forms, and untainted by any trace of subjectivity. An amusing indication of how far this universalism could be taken is that among the hard-line materialists who made up the Cybernetics Group there was one scientist who, without seeming to be too far out of step with his colleagues, saw mind everywhere in nature, or at least in every manifestation of circular organization, whether a whirlpool in a torrent of water, a colony of ants, or an oscillating electric circuit.⁷

It is because the idea of a physics of meaning remains foreign to it that cognitivism is led to make symbolic computation the central pier of the bridge that will enable it, or so it hopes, to bridge the gap that separates mind and meaning from matter. The problems it has encountered arise from the fact that each of the two leaps proves to be a perilous one. On the one hand, the attempt to move from symbolic computation to meaning is open to attacks of the sort made by Searle, as we have seen. With regard to the attempt to move from symbolic computation to the domain of causal physical laws, the difficulty arises from the fact that the semantic and conceptual aspect of computation is not directly given in nature. The cyberneticians did not in principle run into these problems. In the last analysis this was because they took both physics and computation much more seriously than the cognitivists, stripping each one of anything that might call attention to the end to be reached, namely, mind and meaning; but also because they redefined mind and meaning in terms that excluded all reference whatever to psychology and subjectivity. The passage from physics to meaning, thus redefined, is direct, as we have seen. As for the relation between computation and causal physical laws, the cyberneticians had no hesitation in asserting their identity. They held that the laws of physics are computable and therefore that computational models and formalisms are perfectly suited to describing and specifying physical phenomena. Conversely, they were convinced by their theories, whether it was a question of electric circuits or the brain, that logic is embodied in matter, in natural as well as in artificial systems.

Cognitivism resulted from an alliance between cognitive science and the philosophical psychology that is known as philosophy of mind, currently a very active branch of analytic philosophy. Here we have a marriage that never could have occurred with cybernetics, and it is important to understand why. Philosophy of mind set itself the task of rescuing ordinary (or “folk”) psychology by giving it a naturalist and materialist foundation—the only foundation, according to its practitioners, that is capable of conferring scientific legitimacy on any field of research, philosophy included. By “folk psychology” is meant the manner in which people give meaning to the actions and beliefs of their fellow human beings, but also to their own actions and beliefs, explaining them and interpreting them in terms

of *reasons*. These reasons for acting and believing are constituted in turn by the agent's other beliefs, desires, intentions, fears, and so forth, all of them being "mental states" endowed with semantic content. Accordingly, explanation in terms of reasons presupposes and reveals the rationality of individual agents, and therefore possesses a normative component. To "naturalize" this type of explanation requires that it be rooted in a physicalist context in which the ultimate explanations are explanations by causes. It may be thought that naturalizing mind and meaning in this manner is bound to lead to serious errors by confusing the natural and the normative, facts and norms, nature and freedom—in a word, by confusing "is" and "ought." Philosophers of mind believe they have found a way, however, looking to the intermediate computational level postulated by cognitivism, to leap over this obstacle or otherwise to get around it. The computation of the cognitivists, it will be recalled, is symbolic computation. The semantic objects with which it deals are therefore all at hand: they are the mental representations that are supposed to correspond to those beliefs, desires, and so forth, by means of which we interpret the acts of ourselves and others. Thinking amounts, then, to performing computations on these representations. Revising one's beliefs, for example, is a question of inferring a new set of beliefs on the basis of beliefs that one already has, together with an observation that partially disconfirms some of these; rational planning is a question of forming an intention to act on the basis of one's desires and beliefs; and so on. For cognitivism and the philosophical psychology associated with it, conceptual computation of this sort manages to square the circle. It creates an intermediate level between the interpretive level of understanding, where we give meaning to actions and beliefs by means of reasons, and the neurophysiological (ultimately, physical) level, where these actions and beliefs are produced by causal processes. This intermediate level is where "mental causes" are supposed to operate—a hybrid of causes and reasons, or rather reasons treated as causes.

This strategy found itself faced with two sorts of objection, each of which asserted the nonexistence and impossibility of this intermediate level. Cognitivism had no alternative but to fiercely defend itself, for the very heart of its system was under attack. To revert to the image of a bridge, the central pier turned out to be vulnerable, and this pier was none other than the computing machine—more precisely, the computer as a metaphor for thinking. What was called into question was the very wager of cognitive science itself, namely, that the mind could be mechanized and, in this way, made a part of nature. The first objection, due to Wittgenstein's followers, acknowledges the causal level of physical processes, on the one hand, and, on the other, the level of norms, interpretation, under-

standing, and justification, and declares them to be utterly irreducible to each other: no bridge can be thrown up between them. Understanding in terms of reasons constitutes a language game that is incommensurable with that of explanation in terms of causes. The second objection came from “eliminativist” materialists who recognize only a single level, that of physical and physiological causal processes. Folk psychology along with its reasons, which pretend to the same explanatory status as causes, is thus relegated to the trash heap of prescientific illusions. A great part of philosophical debate about cognitive science today may be summarized as a confrontation among these three positions.

It needs to be understood that cybernetics proposed another conceptual approach distinct from the three I have just described. Like eliminative materialism, it banished from its language all talk of reasons, all talk of mental representations having semantic content, and so on. *A fortiori*, then, it recognized no intermediate level of symbolic computation operating on representations. Even so, can cybernetics fairly be characterized as eliminativist? To be sure, it eliminated psychology completely. But what made its approach distinctive was that it did not thereby eliminate the question of meaning.⁸ It redefined meaning by purging it of all traces of subjectivity. Having redefined meaning in this way, it was able to reach it from the far bank of physical causation in a single bound. Since physical causality is computable, and since computation can be implemented in matter, this leap also linked computation with meaning. Had cybernetics succeeded in realizing its ambitions, it would have successfully accomplished the very enterprise—the mechanization of the mind—that cognitivism, proceeding from entirely different assumptions, struggles with today.

In light of this discussion, we can begin to see the outlines emerge of one of the key arguments that this book tries to develop. The major role that analytical philosophy of mind came to play in cognitive science was the result of a historical accident—and surely an unhappy accident at that. How did it happen? Why did it happen? Can the damage be undone? It will be seen that I have only conjectures to offer. I do, however, defend the view that the other great philosophy of mind of the twentieth century, phenomenology, could—and should—have allied itself with cybernetics if, despite the convictions of its founder, Edmund Husserl, it had been interested in providing a natural basis for its doctrines. A whole series of recent works seems to show that this is a fruitful avenue of research.⁹

I have said enough, I trust, to justify the view that all thinking about cognitive science today, about its present state and its history, that does not take into account its origins in cybernetics—as is quite generally the case, given the disrepute into which cybernetics has fallen—will yield only

a very biased view of its current situation and its chances of escaping the impasse into which cognitivism has led it. My chief aim in this book is to provide just such an account. The ideas of cybernetics were good ones. By this I do not mean they were true (in fact, as I say, I am convinced they were not)¹⁰ but that they constituted a coherent model that was perfectly suited to the objective that cognitive science continues to share with cybernetics, which is to say the mechanization of the mind. Those who dedicate themselves to this purpose today may find it useful to immerse themselves once again in these pioneering debates. If any further reason is needed to convince them of this, it would be the following, which is only apparently paradoxical: cybernetics ended in failure. It was a historical failure, one that was all the more bitter as its advertised ambitions were enormous; a conceptual failure, all the less comprehensible in view of the fact that it had marshaled very great intellectual advantages on its side; and, finally, if we consider all that is owed to it and that has gone unacknowledged, it was perhaps an unjust failure as well. However this may be, if cybernetics failed, despite having had so many good ideas, the practitioners of cognitive science today, whose ideas are not necessarily better, would do well to meditate upon the causes and reasons for the failure of cybernetics. The present book ought to a certain extent help them in this.

The Question of Humanism

In addition to the philosophical issues already mentioned, the American reader of this work will perhaps be surprised also to find references both to the work of Heidegger and to the movement of his French followers known as deconstruction (which is to be understood as a shorthand for the “deconstruction of Western metaphysics”). To American academics it may seem altogether incongruous to find associated in the same book—even surreptitiously—von Neumann and Heidegger, cybernetic automata and Lacanian psychoanalysis, Jerry Fodor and Jacques Derrida. The very existence of this sense of incongruity furnishes a fine subject for reflection, as I have already suggested.¹¹ It is therefore necessary at the outset to say a few words about what lies behind this rapprochement of apparently irreconcilable traditions.

In the last analysis, many of the attacks aimed against the materialism of cognitive science are motivated by the desire to provide a defense of humanism. This is not always obvious, for most critics wish to avoid giving the impression of falling into dualism, with its lingering air of religiosity. It required some courage for Thomas Nagel, in criticizing the use made by Jerry Fodor, one of the high priests of cognitivism, of the paradoxical

notion of “tacit knowledge”—a type of knowledge that cognitivism was led to postulate at the intermediate level of computation on mental representations—to recall our traditional way of looking at human beings:

Both knowledge and action are ascribed to individual persons, and they definitely exclude much that the organism can do but that we do not ascribe to the person. Now it may be that these concepts and the distinctions they draw are not theoretically interesting. It may be (although I doubt it) that the idea of a person, with which these other concepts are bound up, is a dying notion, not likely to survive the advances of scientific psychology and neurophysiology.¹²

In raising this disturbing possibility in order then to dismiss it, Nagel poses a classic question: can the idea that we have of the human person, which is to say of ourselves, survive the forward march of scientific discovery? It is a commonplace that from Copernicus to molecular biology, and from Marx to Freud along the way, we have had steadily to abandon our proud view of ourselves as occupying a special place in the universe, and to admit that we are at the mercy of determinisms that leave little room for what we have been accustomed to consider our freedom and our reason. Is not cognitive science now in the process of completing this process of disillusionment and demystification by showing us that just where we believe we sense the workings of a mind, there is only the firing of neural networks, no different in principle than an ordinary electric circuit? The task in which I join with Nagel and others, faced with reductive interpretations of scientific advance of this sort, is to defend the values proper to the human person, or, to put it more bluntly, to defend humanism against the excesses of science and technology.

Heidegger completely inverted this way of posing the problem. For him it was no longer a question of defending humanism but rather of indicting it. As for science and technology, or rather “technoscience” (an expression meant to signify that science is subordinated to the practical ambition of achieving mastery over the world through technology), far from threatening human values, they are on Heidegger’s view the most striking manifestation of them. This dual reversal is so remarkable that it deserves to be considered in some detail, even—or above all—in a book on the place of cybernetics in the history of ideas, for it is precisely cybernetics that found itself to be the principal object of Heidegger’s attack.

In those places where Heideggerian thought has been influential, it became impossible to defend human values against the claims of science. This was particularly true in France, where structuralism—and then poststructuralism—reigned supreme over the intellectual landscape for several decades before taking refuge in the literature departments of American universities. Anchored in the thought of the three great Ger-

manic “masters of suspicion”—Marx, Nietzsche, and Freud—against a common background of Heideggerianism,¹³ the human sciences *à la française* made antihumanism their watchword, loudly celebrating exactly what Thomas Nagel and others dread: the death of man. This unfortunate creature, or rather a certain image that man created of himself, was reproached for being “metaphysical.” With Heidegger, “metaphysics” acquired a new and quite special sense, opposite to its usual meaning. For positivists ever since Comte, the progress of science had been seen as forcing the retreat of metaphysics; for Heidegger, by contrast, technoscience represented the culmination of metaphysics. And the height of metaphysics was nothing other than cybernetics.

Let us try to unravel this tangled skein. For Heidegger, metaphysics is the search for an ultimate foundation for all reality, for a “primary being” in relation to which all other beings find their place and purpose. Where traditional metaphysics (“onto-theology”) had placed God, modern metaphysics substituted man. This is why modern metaphysics is fundamentally humanist, and humanism fundamentally metaphysical. Man is a subject endowed with consciousness and will: his features were described at the dawn of modernity in the philosophy of Descartes and Leibniz. As a conscious being, he is present and transparent to himself; as a willing being, he causes things to happen as he intends. Subjectivity, both as theoretical presence to oneself and as practical mastery over the world, occupies center stage in this scheme—whence the Cartesian promise to make man “master and possessor of nature.” In the metaphysical conception of the world, Heidegger holds, everything that exists is a slave to the purposes of man; everything becomes an object of his will, fashionable as a function of his ends and desires. The value of things depends solely on their capacity to help man realize his essence, which is to achieve mastery over being. It thus becomes clear why technoscience, and cybernetics in particular, may be said to represent the completion of metaphysics. To contemplative thought—thought that poses the question of meaning and of Being, understood as the sudden appearance of things, which escapes all attempts at grasping it—Heidegger opposes “calculating” thought. This latter type is characteristic of all forms of planning that seek to attain ends by taking circumstances into account. Technoscience, insofar as it constructs mathematical models to better establish its mastery over the causal organization of the world,¹⁴ knows only calculating thought. Cybernetics is precisely that which calculates—computes—in order to govern, in the nautical sense (Wiener coined the term from the Greek *χυβερνήτης*, meaning “steersman”):¹⁵ it is indeed the height of metaphysics.

Heidegger anticipated the objection that would be brought against him: “Because we are speaking against *humanism* people fear a defense of the

inhuman and a glorification of barbaric brutality. For what is more *logical* than that for somebody who negates humanism nothing remains but the affirmation of inhumanity?"¹⁶ Heidegger defended himself by attacking. Barbarism is not to be found where one usually looks for it. The true barbarians are the ones who are supposed to be humanists, who, in the name of the dignity that man accords himself, leave behind them a world devastated by technology, a desert in which no one can truly be said to dwell.

Let us for the sake of argument grant the justice of Heidegger's position. At once an additional enigma presents itself. If for him cybernetics really represented the apotheosis of metaphysical humanism, how are we to explain the fact that the human sciences in France, whose postwar development I have just said can be understood only against the background of Heidegger's philosophy, availed themselves of the conceptual toolkit of cybernetics in order to deconstruct the metaphysics of subjectivity? How is it that these sciences, in their utter determination to put man as subject to death, each seeking to outdo the other's radicalism, should have found in cybernetics the weapons for their assaults?

From the beginning of the 1950s—which is to say, from the end of the first cybernetics—through the 1960s and 1970s, when the second cybernetics was investigating theories of self-organization and cognitivism was on the rise, the enterprise of mechanizing the human world underwent a parallel development on each side of the Atlantic. This common destiny was rarely noticed, perhaps because the thought of any similarity seemed almost absurd: whereas cognitive science claimed to be the avant-garde of modern science, structuralism—followed by poststructuralism—covered itself in a pretentious and often incomprehensible philosophical jargon.¹⁷ What is more, it was too tempting to accuse French deconstructionists of a fascination with mathematical concepts and models that they hardly understood.¹⁸ But even if this way of looking at the matter is not entirely unjustified, it only scratches the surface. There were very good reasons, in fact, why the deconstruction of metaphysical humanism found in cybernetics an ally of the first order.

At the beginning of the 1940s, a philosopher of consciousness such as Sartre could write: "The inhuman is merely . . . the mechanical."¹⁹ Structuralists hastened to adopt this definition as their own, while reversing the value assigned to its terms. Doing Heidegger one better, they made a great show of championing the inhuman—which is to say the mechanical.²⁰ Cybernetics, as it happened, was ready to hand, having come along at just the right moment to demystify the voluntary and conscious subject. The will? All its manifestations could apparently be simulated, and therefore duplicated, by a simple negative feedback mechanism. Consciousness? The Cybernetics Group had examined the Freudian unconscious, whose existence was defended by one of its members, Lawrence Kubie, and found it chi-

merical. If Kubie often found himself the butt of his colleagues' jokes, it was not, one suspects, because he was thought to be an enemy of human dignity. It was rather because the postulation of a hidden entity, located in the substructure of a purportedly conscious subject, manifesting itself only through symptoms while yet being endowed with the essential attributes of the subject (intentionality, desires, beliefs, presence to oneself, and so on), seemed to the cyberneticians nothing more than a poor conjuring trick aimed at keeping the structure of subjectivity intact.

It is remarkable, as we shall have occasion to note in some detail, that a few years later the French psychoanalyst Jacques Lacan, along with the anthropologist Claude Lévi-Strauss and the Marxist philosopher Louis Althusser one of the founders of structuralism, should have adopted the same critical attitude toward Freud as cybernetics. The father of psychoanalysis had been led to postulate an improbable "death wish"—"beyond the pleasure principle," as he put it—as if the subject actually desired the very thing that made him suffer, by voluntarily and repeatedly placing himself in situations from which he could only emerge battered and hurt. This compulsion (*Zwang*) to repeat failure Freud called *Wiederholungszwang*, an expression translated by Lacan as "automatisme de répétition," which is to say the *automatism* of repetition. In so doing he replaced the supposed unconscious death wish with the senseless functioning of a machine, the unconscious henceforth being identified with a cybernetic automaton. The alliance of psychoanalysis and cybernetics was neither anecdotal nor fortuitous: it corresponded to a radicalization of the critique of metaphysical humanism.

There was a deeper reason for the encounter between the French *sciences de l'homme* and cybernetics, however. What structuralism sought to conceive—in the anthropology of Lévi-Strauss, for example, and particularly in his study of systems of exchange in traditional societies—was a subjectless cognition, indeed cognition without mental content. Whence the project of making "symbolic thought" a mechanism peculiar not to individual brains but to "unconscious" linguistic structures that automatically operate behind the back, as it were, of unfortunate human "subjects," who are no more than a sort of afterthought. "It thinks" was destined to take the place once and for all of the Cartesian *cogito*. Now cognition without a subject was exactly the unlikely configuration that cybernetics seemed to have succeeded in conceiving. Here again, the encounter between cybernetics and structuralism was in no way accidental. It grew out of a new intellectual necessity whose sudden emergence appears in retrospect as an exceptional moment in the history of ideas.

It is time to come back to our enigma, which now may be formulated as a paradox. Was cybernetics the height of metaphysical humanism, as Heidegger maintained, or was it the height of its deconstruction, as certain of

Heidegger's followers believe? To this question I believe it is necessary to reply that cybernetics was both things at once, and that this is what made it not only the root of cognitive science, which finds itself faced with the same paradox, but also a turning point in the history of human conceptions of humanity. The title I have given to this introduction—the self-mechanized mind—appears to have the form of a self-referential statement, not unlike those strange loops the cyberneticians were so crazy about, especially the cyberneticians of the second phase. But this is only an appearance: the mind that carries out the mechanization and the one that is the object of it are two distinct (albeit closely related) entities, like the two ends of a seesaw, the one rising ever higher in the heavens of metaphysical humanism as the other descends further into the depths of its deconstruction. In mechanizing the mind, in treating it as an artifact, the mind presumes to exercise power over this artifact to a degree that no psychology claiming to be scientific has ever dreamed of attaining. The mind can now hope not only to manipulate this mechanized version of itself at will, but even to reproduce and manufacture it in accordance with its own wishes and intentions. Accordingly, the technologies of the mind, present and future, open up a vast continent upon which man now has to impose norms if he wishes to give them meaning and purpose. The human subject will therefore need to have recourse to a supplementary endowment of will and conscience in order to determine, not what he can do, but what he ought to do—or, rather, what he ought not to do. These new technologies will require a whole ethics to be elaborated, an ethics not less demanding than the one that is slowly being devised today in order to control the rapid development and unforeseen consequences of new biotechnologies. But to speak of ethics, conscience, the will—is this not to speak of the triumph of the subject?

The connection between the mechanization of life and the mechanization of the mind is plain. Even if the Cybernetics Group snubbed biology, to the great displeasure of John von Neumann as we shall see, it was of course a cybernetic metaphor that enabled molecular biology to formulate its central dogma: the genome operates like a computer program. This metaphor is surely not less false than the analogous metaphor that structures the cognitivist paradigm. The theory of biological self-organization, first opposed to the cybernetic paradigm during the Macy Conferences before later being adopted by the second cybernetics as its principal model, furnished then—and still furnishes today—decisive arguments against the legitimacy of identifying DNA with a “genetic program.”²¹ Nonetheless—and this is the crucial point—even though this identification is profoundly illegitimate from both a scientific and a philosophical point of view, its technological consequences have been considerable. Today, as a result, man may be inclined to believe that he is the master of his own

genome. Never, one is tempted to say, has he been so near to realizing the Cartesian promise: he has become—or is close to becoming—the master and possessor of all of nature, up to and including himself.

Must we then salute this as yet another masterpiece of metaphysical humanism? It seems at first altogether astonishing, though after a moment's reflection perfectly comprehensible, that a German philosopher following in the tradition of Nietzsche and Heidegger should have recently come forward, determined to take issue with the liberal humanism of his country's philosophical establishment, and boldly affirmed that the new biotechnologies sound the death knell for the era of humanism.²² Unleashing a debate the like of which is hardly imaginable in any other country, this philosopher ventured to assert: "The domestication of man by man is the great unimagined prospect in the face of which humanism has looked the other way from antiquity until the present day." And to prophesy:

It suffices to clearly understand that the next long periods of history will be periods of choice as far as the [human] species is concerned. Then it will be seen if humanity, or at least its cultural elites, will succeed in establishing effective procedures for self-domestication. . . . It will be necessary, in the future, to forthrightly address the issue and formulate a code governing anthropological technologies. Such a code would modify, a posteriori, the meaning of classical humanism, for it would show that *humanitas* consists not only in the friendship of man with man, but that it also implies . . . , in increasingly obvious ways, that man represents the supreme power for man.

But why should this "superhuman" power of man over himself be seen, in Nietzschean fashion, as representing the death of humanism rather than its apotheosis? For man to be able, as subject, to exercise a power of this sort over himself, it is first necessary that he be reduced to the rank of an object, able to be reshaped to suit any purpose. No raising up can occur without a concomitant lowering, and vice versa.

Let us come back to cybernetics and, beyond that, to cognitive science. We need to consider more closely the paradox that an enterprise that sets itself the task of *naturalizing* the mind should have as its spearhead a discipline that calls itself *artificial* intelligence. To be sure, the desired naturalization proceeds via mechanization. Nothing about this is inconsistent with a conception of the world that treats nature as an immense computational machine. Within this world man is just another machine—no surprise there. But in the name of what, or of whom, will man, thus artificialized, exercise his increased power over himself? In the name of this very blind mechanism with which he is identified? In the name of a meaning that he claims is mere appearance or phenomenon? His will and capacity for choice are now left dangling over the abyss. The attempt to restore mind to the natural world that gave birth to it ends up exiling the

mind from the world and from nature. This paradox is typical of what the sociologist Louis Dumont, in his magisterial study of the genesis of modern individualism, called

the model of modern artificialism in general, the systematic application of an extrinsic, imposed value to the things of the world. Not a value drawn from our belonging to the world, from its harmony and our harmony with it, but a value rooted in our heterogeneity in relation to it: the identification of our will with the will of God (Descartes: man makes himself *master and possessor of nature*). The will thus applied to the world, the end sought, the motive and the profound impulse of the will are [all] foreign. In other words, they are extra-worldly. Extra-worldliness is now concentrated in the individual will.²³

The paradox of the naturalization of the mind attempted by cybernetics, and today by cognitive science, then, is that the mind has been raised up as a demigod in relation to itself.

Many of the criticisms brought against the materialism of cognitive science from the point of view either of a philosophy of consciousness or a defense of humanism miss this paradox. Concentrating their (often justified) attacks on the weaknesses and naiveté of such a mechanist materialism, they fail to see that it invalidates itself by placing the human subject outside of the very world to which he is said to belong.²⁴ What is more, the recent interest shown by cognitive science in what it regards as the “mystery” of consciousness seems bound to accentuate this blindness.²⁵ The dialogue that the present book hopes to inaugurate between the analytic philosophy of mind underlying cognitive science and Continental philosophy, despite the deep-seated prejudices of those on both sides who want them to have nothing to say to each other, finds its justification in precisely this state of affairs.

History of Science vs. History of Ideas

The question of humanism is all the more crucial for a study of the origins of cognitive science since the socioeconomic, political, and cultural context of postwar America in which cybernetics developed is liable to lead the historian of science astray. Owing to circumstances that I shall go on to describe later in the book, cybernetics was obliged from the beginning to ally itself with a movement—a political lobby, actually, operating under the auspices of the Macy Foundation—that sought to assure world peace and universal mental health by means of a bizarre cocktail concocted from psychoanalysis, cultural anthropology, advanced physics, and the new thinking associated with the Cybernetics Group. If only this context is considered, it may seem as though cybernetics was part and parcel of an

effort to place science and technology in the service of human well-being. This interpretation is all the more tempting since one of the most famous participants of the Macy Conferences, and the one who gave cybernetics its name, Norbert Wiener, made himself known to the general public through a series of works that were frankly humanist in tone, through his many stands on behalf of the social responsibility of the scientist, and through various inventions relying on cybernetic advances to help people with disabilities. It is nonetheless my contention that the system of ideas and values embodied by cybernetics can be understood only if one recognizes its purpose as having been fully “antihumanist” in the sense of the preceding discussion.

I have already mentioned in the preface that Steve Heims’s book on the Macy Conferences, *The Cybernetics Group*, appeared before mine. I need to make clear at the outset what I owe to this work and in what respects my approach differs from it.

Heims is a German-born physicist whose family fled Nazism and settled in the United States. He later broke with physics on moral grounds, feeling that the general orientation of research in this field had made it an instrument of inhumanity. He therefore resolved to step back and devote himself to the history of twentieth-century science. He turned his attention first to the careers of two incomparable figures who profoundly influenced the science, technology, and society of their time, before, during, and after the Second World War: John von Neumann and Norbert Wiener. The result was a notable work that took the form of a parallel biography of the two great mathematicians, significantly entitled *John von Neumann and Norbert Wiener: From Mathematics to the Technologies of Life and Death* (1980). Heims’s research led him in turn to undertake a thorough examination of what exactly the cybernetic project had involved, each of his two protagonists having played a crucial role in its creation. Determined to reconstruct as faithfully as possible the circumstances and events of this pioneering period, he decided to write an account of the Macy Conferences, which, despite their very great historical importance, had until then not been the object of systematic study. The result of this second phase of research was *The Cybernetics Group* (1991).

The question naturally arose whether a second work on the Macy Conferences could be justified. After long deliberation, I came to the conclusion that another look at them was in fact needed, for at least two reasons. The first reason is that, despite my great admiration for Heims’s patience and persistence as a historian, I disagree with his approach to the subject. Placing cybernetics in the context of postwar America is, in and of itself, an interesting point of view; but to see it as a reflection of this context, and nothing else, as Heims does, seems to me much more problematic. As I have already indicated, I believe too strongly in the autonomy and the

power of ideas to subscribe to an “externalist” perspective in the sociology of science. In the present instance, such a perspective seems to me to introduce a very serious bias. An essential part of the context of the period, as we have noted, has to do with a movement on behalf of mental health and world peace that was both scientific and humanist—humanist because scientific. Heims takes violent issue with the naive individualist ideology of this movement, which seems to him to obscure the social and political dimension of the problems of the time, without seeing that it concealed a philosophical and scientific project of a wholly different nature. The problem is that he devotes the main part of his book to those members of the Cybernetics Group who belonged to the movement, which gives a highly misleading idea of what cybernetics was actually trying to do. The aim of the cyberneticians was nothing less than to bring about the apotheosis of science by building a science of the mind. It is just this ambition that makes it the parent of cognitive science. That cybernetics should have flourished at this particular moment in history has much less to do, in my view, with the social, political, and ideological atmosphere of the time than with the fact that it was the product of a long evolution in Western thinking about what it is to know. This, at least, is the thesis I defend in the first chapter of the present book. That this thinking should suddenly have crystalized when it did, in the 1940s, was because the shock of the great discoveries in mathematical logic of the preceding decade was needed for it to assume a definite form.

I had felt a similar uneasiness on reading Heims’s first book in which, to oversimplify somewhat, he contrasts a wicked von Neumann with a good Wiener: whereas the former raised the temperature of the Cold War by working to perfect the hydrogen bomb, the latter designed prostheses for the hearing-impaired. The question of the circular relations between science, technology, and society is certainly an important one. But I do not believe it ought to hinder analysis of the internal dynamic that underlies the development of ideas.

The second reason follows directly from the first. Heims is so concerned to denounce capitalism and its ambition of exercising universal control over beings and things that he neglects, or in any case pushes into the background, the question that I have placed at the heart of my inquiry: how far can cybernetics be seen as the womb from which issued what today is called cognitive science? This question, as we shall see, is controversial, and I try to answer it through a series of arguments that jointly constitute a preface to some fuller intellectual history of cognitive science that remains to be written. This has led me frequently to go outside of the period that is the direct object of my inquiry in order to consider the reaction to cybernetics of various figures who came after. I hope I have, for the most part at least, avoided the traps that lay in wait for every historical

essayist; in any case I believe I have not yielded too far to the temptation to take the easy way out by regarding cybernetics as somehow incomplete or deficient by comparison with later research. My resolve in this respect has been strengthened by the feeling that the cybernetic phase exhibited a richness of debate and intuition that cognitive scientists today would do well to contemplate. I do not reproach cybernetics for not having known what was discovered only much later. I reproach cybernetics for not having taken advantage of ideas that were both relevant to its purpose and within its reach, if only it had wanted to take notice of their existence.

My interests are therefore very different than those of Heims, which is why I venture to hope this work may have a place next to his. The reservations that I have expressed with regard to his work do not prevent me from reemphasizing the debt I owe him: lacking the materials he meticulously assembled over a number of years, the foundations of my own inquiry would be far weaker.

What I attempt in this book is an intellectual history that takes the form of an “ecology of mind,” to borrow a phrase from one of the members of the Cybernetics Group, Gregory Bateson. The author of such a history cannot remain wholly in the background. I have already, in the preface, described my motivations and certain of my philosophical positions. Here it is enough simply to say that throughout this book, faced with a particular idea or position, I have not been able to refrain from expressing my enthusiasm and admiration or, depending on the case, my irritation and exasperation. These reactions are unrelated to my opinion of the truth or soundness of the position in question. As I say, I deeply believe that the materialism of cognitive science is wrong. But I much prefer materialists who know what they are about, who seek to work out the most coherent and general theory possible, to ideologues or conceptual tinkerers. Why? Because the only way to prove the falsity of materialism is to give it every benefit of doubt, to allow it to push forward as far as it can, while remaining alert to its missteps, to the obstacles that it encounters, and, ultimately, to the limits it runs up against. Any other method—fixing its boundaries in advance, for example—is bound to fail. Champions of materialism are therefore to be welcomed. I believe I have found such a hero, one who, far more than either Wiener or von Neumann, was the soul of the Macy Conferences: Warren McCulloch. His imposing personality dominates the present work.

Steve Heims, at the end of his book, ranks me among the “happy few” who still today carry on the tradition inaugurated by the first cybernetics. This is a dubious honor, which I am not sure I deserve. The verdict I deliver on the events described in this book is rather harsh. It is offered as the history of a failure—a grandiose failure, if you like, and in any case a productive one full of lessons to be learned; but a failure nonetheless, if

what was actually achieved in the end is set against what was hoped for at the beginning. I said a moment ago that I have taken care to avoid the traps of retrospective history. Instead I have dared to practice a new and rather hazardous genre: counterfactual history. For cognitive science in its earliest stages missed so many important rendezvous, so many important opportunities, that the historian—still less the philosopher—cannot resist asking, again and again, “What would have happened if only . . . ?” A pointless exercise, perhaps—too facile, certainly—but one that nonetheless conveys this philosopher’s abiding sense of frustration.