

The Mechanical Mind in History

edited by Philip Husbands, Owen Holland, and Michael Wheeler

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

1. Charles Babbage: The Dawn of the Computer

For information about special quantity discounts, please email orders@mit.edu

2. Alan Turing: A Grandfather of Artificial Intelligence

Margaret A. Boden

This book was set in Stone Serif and Stone Sans on 382 by Ascot Typesetters, Hong Kong. Printed and bound in the United States of America.

3. Alan Turing: A Grandfather of Artificial Intelligence

Donald Michie

Library of Congress Cataloging-in-Publication Data

4. The Mechanical Mind in History

The mechanical mind in history / edited by Philip Husbands, Owen Holland, and Michael Wheeler.

5. The Rube Club: A History of Rube Goldberg's Inventions

Philip Husbands, Owen Holland, and Michael Wheeler

6. From Mechanisms of Adaptation to Artificial Intelligence

Includes bibliographical references and index.

7. Artificial Intelligence—Philosophy. I. Husbands, Philip. II. Holland, Owen. III. Wheeler, Michael, 1960-

8. Gordon Pask and His Maverick Machines

Gordon Pask and His Maverick Machines

9. Sandbag Dreaming

Andy Becklet

10. Toward the Synthetic Method: Symbolic Information Processing

11. Organizing Systems in Early Artificial Intelligence Modeling

A Bradford Book

The MIT Press

Cambridge, Massachusetts

London, England

Contents

Preface	vii
1 Introduction: The Mechanical Mind	1
Philip Husbands, Michael Wheeler, and Owen Holland	
2 Charles Babbage and the Emergence of Automated Reason	19
Seth Bullock	
3 D'Arcy Thompson: A Grandfather of A-Life	41
Margaret A. Boden	
4 Alan Turing's Mind Machines	61
Donald Michie	
5 What Did Alan Turing Mean by "Machine"?	75
Andrew Hodges	
6 The Ratio Club: A Hub of British Cybernetics	91
Philip Husbands and Owen Holland	
7 From Mechanisms of Adaptation to Intelligence Amplifiers: The Philosophy of W. Ross Ashby	149
Peter M. Asaro	
8 Gordon Pask and His Maverick Machines	185
Jon Bird and Ezequiel Di Paolo	
9 Santiago Dreaming	213
Andy Beckett	
10 Steps Toward the Synthetic Method: Symbolic Information Processing and Self-Organizing Systems in Early Artificial Intelligence Modeling	219
Roberto Cordeschi	

7 From Mechanisms of Adaptation to Intelligence Amplifiers: The Philosophy of W. Ross Ashby

Peter M. Asaro

During the last few years it has become apparent that the concept of “machine” must be very greatly extended if it is to include the most modern developments. Especially is this true if we are studying the brain and attempting to identify the type of mechanism that is responsible for the brain’s outstanding powers of thought and action. It has become apparent that when we used to doubt whether the brain could be a machine, our doubts were due chiefly to the fact that by “machine” we understood some mechanism of very simple type. Familiar with the bicycle and the typewriter, we were in great danger of taking them as the type of all machines. The last decade, however, has corrected this error. It has taught us how restricted our outlook used to be; for it developed mechanisms that far transcended the utmost that had been thought possible, and taught us that “mechanism” was still far from exhausted in its possibilities. Today we know only that the possibilities extend beyond our farthest vision.

—W. Ross Ashby (1951, p. 1).

The idea that intelligence could be imitated by machines has appeared in numerous forms and places in history. Yet it was in the twentieth century, in Europe and North America, that these metaphorical ideas were transformed into scientific theories and technological artifacts. Among the numerous scientists who pursued mechanistic theories of intelligence in the last century, W. Ross Ashby (1903–1972) stands out as a particularly unique and interesting figure. A medical doctor and psychiatrist by training, Ashby approached the brain as being first and foremost an organ of the body. Like other organs the brain had specific biological functions to perform. Ashby further believed that through a thoughtful analysis of those functions, a quantitatively rigorous analysis of the brain’s mechanisms could be devised. It was his single-minded dedication to this basic idea that motivated his research into the mechanisms of intelligence for more than forty years. By always insisting upon sticking to the naturalistic

functions of the brain, and to quantitative methods, Ashby was led to a number of startling and unique insights into the nature of intelligence that remain influential.

In this chapter I seek to sketch an intellectual portrait of Ashby's thought from his earliest work on the mechanisms of intelligence in 1940 through the birth of what is now called Artificial Intelligence (AI), around 1956, and to the end of Ashby's career in 1972. This period of Ashby's intellectual development is particularly interesting in his attempts to grasp the basic behaviors of the brain through the use of mechanical concepts. It is unique in the way that Ashby used rather sophisticated mechanical concepts, such as equilibrium and amplification, which were not particularly favored by other researchers. And moreover, he used these concepts not merely metaphorically, but also imported their associated mathematical formulations as a basis for quantifying intelligent behavior. As a result of this, we can see in Ashby's work both great insight and a truly original approach to the mechanisms of intelligence.

Ashby's professional career, beginning in 1928 and lasting until his death, is itself a remarkable tale that merits further research. He was the author of two enormously influential books in the early history of cybernetics, *Design for a Brain* (1952c) and *An Introduction to Cybernetics* (1956b).¹ Between his written contributions and his participation in the scientific community of cybernetics and its conferences and meetings, Ashby is considered to be one of the pioneers, or even cofounders, of cybernetics, which in turn gave rise to AI.

Our primary concern, however, will be with the central tenets of Ashby's thought. In particular we seek to discover the problems that motivated his thought, the conceptual form that he gave to those specific problems, and how their resolution resulted in a new mechanistic understanding of the brain and intelligence. This recounting of Ashby's mental philosophy will proceed in a roughly chronological fashion. We shall begin by examining his earliest published works on adaptation and equilibrium, and the conceptual structure of his notions of the mechanisms of control in biological systems. In particular we will examine his conceptions of mechanism, equilibrium, stability, and the role of breakdown in achieving equilibrium. We shall then proceed to his work on refining the concept of "intelligence," on the possibility of the mechanical augmentation and amplification of human intelligence, and on how machines might be built that surpass human understanding in their capabilities. I conclude with a consideration of the significance of his philosophy, and its role in cybernetic thought.



Figure 7.1

Ashby in front of his house, Westons, in 1960. Used with permission of the Trustees of the Estate of W. Ross Ashby.

The Mechanism of Adaptation

Given that Ashby was trained in medical psychiatry, and that his early work focused on neurological disorders from a strongly medical and physiological perspective, it might seem curious that he should come to be one of the leading proponents of a mechanical perspective on the mind. Mechanics has had a long and successful scientific history, and certainly scientists and philosophers before him had submitted that the brain, and perhaps also the mind, were in some sense machine-like. Roberto Cordeschi (2002) has carefully illustrated how a group of psychologists were arguing about possible mechanisms that could achieve mental capabilities, and were seeking to give a purely mechanistic explanation of mental capacities in the early decades of the twentieth century. Yet these scientific debates dwelled on the proper ways to separate out the mechanistic from the metaphysical aspects of psychology—consciousness, voluntary actions, and the spiritual aspects of mind. These scientists did propose specific types of mechanisms, such as Jacques Loeb's (1900) orientation mechanisms, and also built electronic automata to demonstrate these principles, such as John Hammond Jr. and Benjamin Miessner's (1915) phototropic robot (Miessner 1916). While these sorts of behaviors were interesting, for Ashby they were not sufficient to demonstrate that intelligence itself was mechanistic. Ashby knew that a mechanistic approach to the mind would have to deal with the most complex behaviors as well as the simplest, and do so with a single explanatory framework. It was with this goal in mind that he elaborated on the mechanistic nature of adaptation, as a route from simple physiology to complex forms of learning.

Another aspect of Ashby's work, shared with the pre-cybernetic and cybernetic mechanists, was that the development of theories of the brain and behavior went hand in hand with the development of technologies that exploited these theories in novel artefacts. Ashby summarized his own intellectual career in 1967 by saying (1967, p. 20):

Since opening my first note-book on the subject in 1928, I have worked to increase our understanding of the mechanistic aspect of "intelligence," partly to obtain a better insight into the processes of the living brain, partly to bring the same processes into action synthetically.

In many ways the construction of synthetic brains was integral to the theorization of the living brain. Cordeschi (2002) has called this approach the "synthetic method," and it continues in many areas of AI and robotics.² Although this essay focuses on the theoretical development of Ashby's

thought, there is a deep technological aspect to that development and the machines Ashby built are worthy of consideration in their own right (Asaro 2006).

To understand how the key aspects of the transformation of psychological concepts to mechanical explanations took place in Ashby's thought, we must look at the unique way in which he reconceptualized the observed behavior of thinking creatures as being equivalent to the mechanical processes of physical devices. Ashby's views on these matters warrant careful consideration insofar as they do not fall easily into the categories employed by contemporary philosophers of mind, such as reductive materialism or straightforward functionalism. Ashby (1952e) did see his objective as being to provide a physical explanation of the mind (p. 408; emphasis in all excerpts is as in the original except where noted):

The invasion of psychology by cybernetics is making us realize that the ordinary concepts of psychology must be reformulated in the language of physics if a *physical* explanation of the ordinary psychological phenomena is to become possible. Some psychological concepts can be re-formulated more or less easily, but others are much more difficult, and the investigator must have a deep insight if the physical reality behind the psychological phenomena is to be perceived.

But his views on this matter are rather more complex than merely attempting to reduce mental processes to physical or physiological processes in the brain. As he expressed in a review of J. C. Eccles's *The Neurophysiological Basis of Mind* (Ashby 1954, p. 511):

The last two chapters, however—those on the cortex and its highest functions—fall off sadly, as so often happens when those who have spent much time studying the *minutiae* of the nervous system begin to consider its action as a whole; yet it is difficult to see, while present-day neurophysiology is limited to the study of the finest details in an organism carefully isolated from its environment, how the neurophysiologist's account could have been improved. The last two chapters, in fact, show only too clearly how ill adapted classical neurophysiology is to undertake the study of the brain's highest functions. At the moment it is far too concerned with details, and its technical resources are leading it only into the ever smaller. As a result, the neurophysiologist who starts to examine the highest functions is like a microscopist who, hearing there are galaxies to be looked at, has no better resource than to point his microscope at the sky. He must not be surprised if he sees only a blur.

Ashby recognizes that the instruments of investigation shape what one finds, and the question is what instruments to use to study the brain. Like other scientists who were trying to draw similar conclusions about the physical basis of mentality at the time, Ashby did believe that mental and

psychological processes were essentially physical and chemical processes, but he argued that this did not mean that they could be explained and understood by simply appealing to some deeper or more fundamental level of analysis, such as physiology, in the quote. He believed that the methodology of physical analysis could be applied to mental states directly, the way statistical mechanics could be applied to a volume of gas to describe its behavior without being concerned with the motions of the individual molecules within the gas in order to characterize the relationships between pressure, volume, temperature, and so forth. Thus, Ashby sought to apply mechanistic analysis to the gross holistic organization of behavior directly, not merely to low-level processes, and to thereby demonstrate the general mechanisms by which the brain could achieve mental performances.

The first step in this conceptual move was not a purely metaphysical argument, though its conclusion had profound metaphysical implications. It was primarily an epistemological argument by analogy. Instead of considering the metaphysical arguments directly, he took an epistemological approach which sought to explain the mental phenomena of “adaptation” by an analogy to a physical mechanical process of “equilibrium.” This approach is epistemological insofar as it attempts to show that we can *know or understand* the mind the same way we *understand* mechanical processes—by virtue of the analogy made between them. This is in contrast to others, who pursued a metaphysical argument that the mind must submit to mechanistic explanation because it was necessarily made up of the obviously physical brain—though Ashby also believed this, indeed took it for granted. His particular argument by analogy in fact appeals to the metaphysical necessity of equilibrium, but rather than argue that adaptation is reducible to this concept, shows that it is equivalent, and hence can be *analyzed and studied in the same manner as mechanical processes but independent of its specific material composition*. And so, it is *how* one comes to know a thing that is primary to the argument, and not its “essence.”

The central argument of Ashby’s mechanistic approach first appears in “Adaptation and Equilibrium” (1940). The title discloses the two concepts that he argues are analogous. In its final formulation, the analogy he argued for was that adaptive behavior, such as when a kitten learns to avoid the hot embers from a fire, was equivalent to the behavior of a system in equilibrium. In establishing this analogy, he shows that the biological phenomena of adaptive behavior can be described with the language and mathematical rigor of physical systems in states of equilibrium. In his own summary (p. 483):

Animal and human behavior shows many features. Among them is the peculiar phenomenon of “adaptiveness.” Although this fact is easily recognized in any given case, yet it is difficult to define with precision. It is suggested here that adaptive behavior may be identical with the behavior of a system in stable equilibrium, and that this latter concept may, with advantage, be substituted for the former. The advantages of this latter concept are that (1) it is purely objective, (2) it avoids all metaphysical complications of “purpose,” (3) it is precise in its definition, and (4) it lends itself immediately to quantitative studies.³

Thus Ashby suggests that a well-understood mechanical concept, carrying with it an extensive set of mathematical tools, ought be substituted for the vague conception of adaptive behavior in common usage. This passage also makes clear that Ashby’s motivation in seeking a mechanistic explanation of mental phenomena is to provide a new basis for scientific study, and to sidestep rather than resolve any outstanding philosophical problems. It is also apparent that he was aware of the metaphysical issues surrounding the mind and believed that by conceiving of adaptation as equilibrium in this way one could avoid them.

The first half of the analogy depends upon establishing the importance of adaptive behavior in living and thinking things. Ashby begins by arguing that a peculiar feature of living organisms is their adaptive behavior. While definitions of life might variously include such requirements as motive, vegetative, or reproductive capacities, essential to this argument was the notion that the capacity for adaptation is necessary, and possibly sufficient, for something to be a living organism. In his second paper on the subject, “The Physical Origin of Adaptation by Trial and Error” (1945), Ashby elaborated on the role of adaptation in biological organisms, and to this end quoted various biologists, including Jennings (p. 14, quoting Jennings 1915):

Organisms do those things that advance their welfare. If the environment changes, the organism changes to meet the new conditions. . . . If the mammal is cooled from without, it heats from within, maintaining the temperature that is to its advantage. . . . In innumerable details it does those things that are good for it.

It is important to note that Ashby did not restrict his conception of adaptation to the Darwinian notion of adaptation by natural selection, though he certainly considered this to be a profoundly important form of adaptation, as his later writings make clear. Adaptation is then quickly extended from the physiological reactions of whole species to include also the notion of a behavioral response to a novel stimulus by an individual animal—the groundwork for a bridge between biology and behavioral psychology—and

further generalized to include any observable behavior at all. In Ashby's favorite example, the kitten will not at first avoid the glowing embers from a fire, will burn its paw, and will thereafter avoid the fire; the resulting observed behavior is "adapted" insofar as it was the result of the kitten's individual experience of the world.⁴

The other half of the analogy, equilibrium, was seen to provide a rigorous set of analytical tools for thinking about the mind by importing the mathematical theory of mechanisms. Equilibrium is initially defined as a metaphysical necessity (Ashby 1940, p. 482):

Finally, there is one point of fundamental importance which must be grasped. It is that stable equilibrium is necessary for *existence*, and that systems in unstable equilibrium inevitably destroy themselves. Consequently, if we find that a system *persists*, in spite of the usual small disturbances which affect every physical body, then we may draw the conclusion with absolute certainty that the system must be in stable equilibrium. This may sound dogmatic, but I can see no escape from this deduction.

Ashby later (1945) employed the simpler definition of the physicist Hendrik Lorentz (1927): "By a state of equilibrium of a system we mean a state in which it can persist permanently" (p. 15). Since many equilibrium states are precarious and unlikely, Ashby further qualifies this by accepting the definition of a "stable" equilibrium as one in which a system will return to the equilibrium state even when some of its variables are disturbed slightly. For example, a cube resting on a table is in a stable equilibrium since it will return to the same state if tilted slightly and released. By contrast, though it might be possible to balance a cone on its point, under the slightest disturbance it will *not* return to the balanced state but will fall into a remote state and thus is in an odd sort of equilibrium if so balanced—an "unstable" equilibrium. A sphere resting on a table represents a "neutral" equilibrium, which is stable at many adjacent states and can be moved freely and smoothly between those states.⁵ He clarifies the concept's meaning (Ashby 1940, pp. 479, 483):

We must notice some minor points at this stage. Firstly, we notice that "stable equilibrium" does *not* mean immobility. A body, e.g. a pendulum swinging, may vary considerably and yet be in stable equilibrium the whole time. Secondly, we note that the concept of "equilibrium" is essentially a dynamic one. If we just look at the three bodies [cube, cone, and sphere] on our table and do nothing with them the concept of equilibrium can hardly be said to have any particular meaning. It is only when we disturb the bodies and observe their subsequent reactions that the concept develops its full meaning. . . .

The question of whether adaptiveness is *always* equivalent to "stable equilibrium" is difficult. First we must study the nature of "adaptiveness" a little closer.

We note that in all cases adaptiveness is shown only in relation to some specific situation: an animal in a void can show neither good nor bad adaptation. Further, it is clear that this situation or environment must affect the animal in some manner, i.e. must change it, since otherwise the animal is just receiving the stimulus without responding to it. This means that we are dealing with a circuit, for we have, first: environment has an effect on the animal, and then: the animal has some effect on the environment. The concept of adaptive behavior deals with the relationship between the two effects. It becomes meaningless if we try to remove one of the effects.

These points are by no means minor, but reflect Ashby's insistence on explaining the dynamic processes of observable phenomena, and how this can be done in terms of mechanisms seeking equilibrium.

The emphasis on "behavior" here, and throughout Ashby's work, is probably best read not as a commitment to, or sympathy for, behaviorism, but as an insistence on the epistemological limitations of science to observable phenomena. "Adaptation," like other scientific concepts, is nothing more than a set of observed reactions of various systems under different conditions. Those conditions are crucial insofar as the environment provides the context for the actions and reactions—the behavior—of a system, a necessary link in the chain of cause and effect. "Observation" is also crucial here, as it is throughout cybernetics, as the basis for determining the system and phenomena in question—both are meaningless in the absence of an observer. This is most likely an inheritance from positivism, which Ashby's approach shared to some extent with behaviorism in its insistence on "observable behaviors" in the form of responses in conditioned response. Although Ashby drew on behaviorist methodology, he went beyond its theory to posit the mechanism that controlled and extended behaviors. Pavlovian conditioning reinforced existing behaviors, and explained responses to stimuli based on this type of conditioning, but made no attempt to explain the mechanisms that supported this kind of conditioning.

Mechanical theory was of particular interest to Ashby by virtue of its potential for supplying a mathematical basis for psychology. A mathematical model of a state-determined mechanical system, such as those used by engineers at the time, involves several parameters divided into variables and constants in a set of equations or functions. When such a model is of a linear dynamical system, the values of the variables at one time determine the values at future times in a deterministic fashion—the functions generate the values for the next time-step from the values at the current time-step. The values of the variables in such a system may eventually stop changing. For example, if we were to observe the value of the angular

displacement of a pendulum—how far it is from pointing straight down—that value would appear to grow and shrink and grow a little less with each swing until it eventually settled down to zero. An equilibrium in these systems is an assignment of values to the variables such that the variables will not change in future time-steps under the rules governing the system, such as when the pendulum rests pointing straight down. If a particular model does not have an equilibrium state, the variables will continue changing endlessly, typically with their values going to extreme limits. Such systems, Ashby argues, are not often found in nature—he can think only of a comet being hurled into deep space, never to return. Most of the systems found in nature, as well as human-made machines, have equilibria in which the variables settle to constant or cyclically repetitive values.

In fact, when an actual machine does not arrive at an equilibrium, it exhibits an intriguing phenomenon—it *breaks* (Ashby 1945, p. 17):

What happens to machines, as defined above, in time? The first point is that, in practice, they all arrive sooner or later at some equilibrium (in the general sense defined above). Thus, suppose we start with a great number of haphazardly assembled machines which are given random configurations and then started. Those which are tending towards equilibrium states will arrive at them *and will then stop there*. But what of the others, some of whose variables are increasing indefinitely? In practice the result is almost invariable—something breaks. Thus, quicker movements in a machine lead in the end to mechanical breaks; increasing electric currents or potentials lead inevitably to the fusing of wires or the break-down of insulation; increasing pressures lead to bursts; increasing temperatures lead to structures melting; even in chemical dynamics, increasing concentrations sooner or later meet saturation.

A break is unlike the normal changes in a dynamic machine in an important way. A break is a change in the organization of a system. In changing its organization, the machine ceases to be the machine it was and becomes a new machine. In the mathematical theory of mechanisms, the equations or functions that previously defined the system no longer hold true. To describe the change mathematically we must either define a new system of equations or must have previously defined a set of equations containing constants whose values can represent the current and alternate organizations of the machine. When the machine “breaks,” those constants change and consequently the relationships between the variables of the system suddenly become different. And while the variables in a system can change either in discrete steps or continuously, a break, or change in the constants, is necessarily a discontinuous change from one distinct organization to another distinct organization—in other words, a shift from one set of equations to another set of equations.

Given this understanding of equilibrium and the dynamics of machines, the analogy to adaptation becomes clear (Ashby 1945, p. 17):

We may state this principle in the form: dynamic systems stop breaking when, and only when, they reach a state of equilibrium. And since a “break” is a change of organization, the principle may be restated in the more important form: *all dynamic systems change their internal organizations spontaneously until they arrive at some state of equilibrium.*

The process of breaking continues indefinitely as long as the variables describing the system continue to exceed tolerable limits on their values—that is, until the variables can be kept within certain limits. The instances of unbounded variables in nature, like the comet, are quite rare. By then applying this understanding to biological organisms, he argues that the organism adapts to its environment by successive trials of internal reorganization until it finds an equilibrium in which its physiological needs are met. In later writings, Ashby (1952a, c) will stress the importance of certain “essential variables,” which the organism must maintain within certain limits in order to stay alive, such as body temperature, blood sugar level, and so forth. In its psychological formulation, the thinking system behaves so as to seek and approach a “goal,” defined as a set of desired values over certain variables. The organism thus seeks to find an equilibrium of a special kind, one in which essential variables are kept within their safe and vital limits, or in which a goal is satisfied.

What seems perhaps most curious in this conceptual transformation is the productive power placed in breakdowns. Generally, a breakdown is seen as undesirable, something to be avoided, and the mark of a bad machine. Here it has become the supreme virtue of living machines: the creative drive, the power to generate alternative organizations in order to adapt to the environment. This result is in part due to the rigid structures of mathematics: it is easy to represent change in variables, but a change *in the relationships between variables* cannot be as easily expressed. In order to describe a machine that changes its dynamics, it is necessary to switch from one set of functions to another. Ultimately, Ashby would cease using the language of “breakdowns” and replace it with the language of “step-functions,” a mathematical formulation that broadened the representation of a system to include its possible organizations and the discontinuous transitions between those organizations.

A similar tension is reflected also in the seeming banality of equilibrium—a system in equilibrium just stops, every dead thing and piece of inert matter is in a state of equilibrium. How can equilibrium be the

ultimate goal of life when it implies a kind of stasis? What makes one kind of equilibrium indicative of life, is that it is *dynamic and is not uniform over the total system*. The living system can maintain some desired portion of its organization in equilibrium, the essential variables, even as the rest of the system changes dynamically in response to disturbances that threaten to destroy that desired equilibrium. For Ashby, this involved developing his conception of “ultrastability”—the power of a system to always find a suitable equilibrium despite changes in its environmental conditions. That is, the organism achieves a certain kind of stability for a few vital variables such as blood-sugar level, by varying other variables that it controls, sometimes wildly, as when an animal searches for food to maintain its blood-sugar levels.

The idea of equating adaptation and equilibrium appears to be unique to Ashby, though it bears strong similarities to ideas such as “negative feedback,” which were being developed by other cyberneticians at the time. Ashby continued to cite and restate this analogy and argument throughout his career and used it as the basis of his first book, *Design for a Brain* (1952c); he never changed it significantly. Once it was published, he appears to have focused his energies on promoting the idea in various ways, including explicating its relationship to the ideas of other cyberneticians, including “negative feedback,” and finding new expressions of the idea in his writings and in working machines. We now turn to the most notorious of these machines.

The Homeostat, completed in 1948, is a fascinating machine for several reasons. Most obvious is that it is a machine with an odd sort of purpose. It does not “do” anything in the sense that a machine generally serves some useful human purpose; unlike a bicycle or typewriter, it has no real practical application. On the other hand, it has its own “purpose” in the purest sense given by cybernetics: its equilibrium-seeking behavior is goal-oriented and controlled by negative feedback and so it is a teleological mechanism. This means that the machine itself has a goal, as revealed by its behavior, which may or may not have anything to do with the goals of its designer, a distinction that was to be further elaborated in Ashby’s philosophy.

Most interesting, perhaps, is its role as a scientific model (Asaro 2006). It stands as a working physical simulation of Ashby’s theory of mental adaptation. As a simulation it offers a powerful illustration of his conception of adaptive behavior in all kinds of systems, and in this regard its isomorphic correspondence to elements of his abstract theory are crucial. To see these correspondences, a brief description of the device is helpful.

The classic setup of the Homeostat consisted of four independent units, each one connected directly to each of the other three through circuits whose resistance could be controlled by either a preset switch or a randomizing circuit, called a “uniselector.” They could “adapt” to one another by adjusting the resistances in the circuits that connected them, provided that the uniselector was engaged instead of the preset switches. Each unit featured a trough of water on top that contained an electrical field gradient and that had a metal needle dipping into it. By virtue of its connection to the current from the other units via the resistors and uniselectors, this needle acted as an indicator of the state of the unit: being in the middle of the trough represented a “stable” position, and being at either end of the trough represented an unstable position. Due to a relay that involved the position of the needle, whenever the needle was outside a central position in the trough it would send a charge to a capacitor. When the capacitor reached a predetermined charge level it would discharge into the uniselector, causing it to switch to a new random resistance in the circuit. These were only pseudo-random, however, as the resistances were derived from a table of random numbers and hard-wired into the uniselector, which stepped through them sequentially (see figure 6.3, p. ■■■, for a photograph of the device).

The correspondence between the Homeostat and Ashby’s theory of mechanistic adaptation rests on an isomorphism between “random variations” and the operation of the uniselector circuit elements; between “acceptable values for essential variables” and the relay controlling the energizing capacitor for the uniselectors; between “equilibrium” and the visible needle resting in the middle of the trough; and between the wildly behaving needles of a machine out of control and a system that continues to “break” up its internal organization through step-functions until it finds equilibrium.

In a later paper, “Simulation of a Brain,” Ashby (1962) discusses the objectives of modeling and simulation directly. In that paper he defines a model formally as a system that stands in relation to another system by virtue of an explicit mapping between sets of elements. He asserts that physical as well as mathematical and symbolic forms can stand in such relationships. He also insists that the value of the formal definition is that it provides a quantitative measure of the closeness of a model to the original system by virtue of the number of relationships shared among the members of the two sets. Given this definition of a model, he argues that there are three virtues to simulations, as physical models, which contribute to scientific progress. The first is their *vividness*: to clearly express a concept

in an easily graspable form. The second is their function as an *archive*: to stand as a repository of built-up knowledge that might be too vast and complex to be written out or grasped all at once by an individual. The final virtue of simulations is their capacity to facilitate *deduction and exploration*: to resolve disputes, disprove hypotheses, and provide a basis for scientific inquiry into areas that, without simulations, would otherwise remain speculative (Ashby 1962, pp. 461–64). He offers the Homeostat as an example of a simulation useful in scientific education for demonstrating that goal-seeking behavior, as a trial-and-error search for equilibrium, presents a fundamentally different kind of mechanical process—negative feedback with step-functions—and opens up new vistas of possibility for what machines might be capable of doing. I have argued elsewhere (Asaro 2006) that working brain models such as the Homeostat also served an important role in mediating between theories of behavior and physiological theories of neurons in the development of the mechanistic theory of the mind.

Designs for Intelligence

With the analogy between adaptation and equilibrium firmly in place, Ashby turned his attention to demonstrating the significance and potential applications of this new insight. His effort consisted of two distinct parts: the development of other simulations, such as the Dispersive And Multi-stable System (DAMS) made of thermionic valves and neon light tubes (Ashby 1951), in order to demonstrate his ideas in more tangible forms; and the continuing articulation of a clear and compelling rhetorical framework for discussing the problems of designing intelligent machines. The machines Ashby developed are deserving of further study as technological artifacts built on unique principles of design, but a discussion of these would take us to remote regions of his mental philosophy, whereas we are concerned only with its central features. In the following sections, we will consider the further development of his theoretical views. We shall begin by looking at Ashby's formal articulation of a "problem" that his mechanism of adaptation could "solve," and then to how this problem-solving mechanism could be generalized to solving more significant and compelling problems. In so doing we shall examine his definition of intelligence and how it could be fully mechanized. Throughout these efforts, Ashby sought to motivate and inspire the belief that a revolution had occurred in our understanding of machines, and that the mechanism of adaptation might ultimately result in machines capable of impressive and even super-human performances.

The Problem of the Mechanical Chess Player

While satisfied with the soundness of his argument for the possibility of an adaptive mechanism, Ashby felt compelled to demonstrate the full significance and implications of this possibility to an audience beyond the handful of psychiatrists and cyberneticians with whom he had contact. To do this, he developed a clear and compelling problem through which audiences could grasp this significance. The example he elaborated on was the “Problem of the Mechanical Chess Player,” which he credited to his experiences in casual conversations, most likely with the members of the Ratio Club, such as Alan Turing, who were very interested in the mathematical problems of chess play. Ashby took the problem in a different direction than Turing and subsequent AI researchers did, and used this as an imaginative, and thus compelling, example of the basic problem of the very possibility of mechanized thought, which could be formalized using the analytical apparatus borrowed from mechanical theory. The rhetorical development of the problem of the mechanical chess player is interesting because it starts by raising some fundamental issues of metaphysics, but once properly formulated as a technical problem, it could be decisively resolved by the demonstrated performance of a working machine. Just how this was achieved we shall now see.

The metaphysical problem of the mechanical chess player was how (or in its weaker form, *whether*) it could be possible to design a machine that has a greater range or skill in performance than what its designer had provided for it by its design—in other words, whether a mechanical chess player can *outplay its designer*. As Ashby (1952d) posed the question in the Ninth Josiah Macy Jr. Foundation Conference on Cybernetics (p. 151):

The question I want to discuss is whether a mechanical chess player can outplay its designer. I don't say “beat” its designer; I say “outplay.” I want to set aside all mechanical brains that beat their designer by sheer brute power of analysis. If the designer is a mediocre player, who can see only three moves ahead, let the machine be restricted until it, too, can see only three moves ahead. I want to consider the machine that wins by developing a deeper strategy than its designer can provide. Let us assume that the machine cannot analyze the position right out and that it must make judgements. The problem, then, becomes that the machine must form its own criteria for judgement, and, if it is to beat its designer, it must form better judgements than the designer can put into it. Is this possible? Can we build such a machine?

While Ashby chose to formulate the problem as whether a machine can outplay its designer, it seems less confusing to me to formulate it as whether a machine can outplay its design, that is, whether it can do “better”

than it was designed to, rather than to say that it can actually defeat the person who designed the machine. In short, Ashby was concerned with the ability of a machine, in this case a chess-playing machine, to acquire knowledge and skill beyond the knowledge and skill built into it.

Ashby hoped to show this by arguing that a mechanism utilizing a source of disorganized information, though one containing a greater variety of possibilities than the designer could enumerate, could in principle achieve better strategies than its designer. Because a generator of random moves could produce novel moves that no known specific or general rule of chess would suggest, there was a possibility of finding a “supermove” that would not otherwise be found and so could not have been built into the machine. Therefore, as long as a system was designed so as to allow the input of such random possibilities, and designed with the ability to select among those possibilities, it might be possible for it to find moves and strategies far better than any its designer could have provided.

This particular formulation in fact caused some confusion at the Macy Conference. In the ensuing discussion of it, Julian Bigelow challenged the distinction Ashby attempted to make between analysis and strategic judgment (Ashby 1952d, pp. 152–54).⁶ For Bigelow, the ability to construct strategies was itself already a kind of analysis. He argued that limiting the analysis of the system to looking only three moves ahead necessarily put a limitation on the number of strategies that could be considered. He also rejected the notion that adding random noise could add any information to the chess-playing system at all—for him information necessarily had to have analytical import and random noise had none. To provide a resolution of this confusion and a better understanding of the role of this problem in thinking machines more generally, we must first clarify Ashby’s conception of “design” and “designer,” as well as the formal articulation he gave to the problem.

Ashby saw the issue as a fundamentally philosophical problem of agency having its roots deep within the tradition of European thought. He offered, as different formulations of the same problem, the following examples from that tradition: “Descartes declared that there must be at least as much reality and perfection in the cause as in the effect. Kant (*General History of Nature*, 1755) asked, ‘How can work full of design build itself up without a design and without a builder?’” (Ashby 1952b, p. 44). Descartes’s dictum, of course, maintains that an effect cannot have more perfection than its cause, and thus a designed system cannot be superior to its designer.⁷ If true, the implication of this dictum is that a machine, being capable only of what its design has provided for it, can never be “better” than

that design, and thus cannot improve on it. But Ashby believed that he had already shown how a mechanism could be capable of adaptation—a kind of improvement relative to environmental conditions. He thus saw it as essential to prove that Descartes was wrong, and saw that the proof would require a more rigorous formal presentation.

The crux of the problem lay in the proper definition of “design.” For a proof, it was necessary to provide a formal definition that could show clearly and quantitatively exactly what was contained in the “design” provided by a designer, such that this could be compared to the quantity of the “design” demonstrated in the performance of the machine. He derived these measures using the information theory of Claude E. Shannon (1948). The quantities measured in the “design” and in the machine would be information, and if a machine could be shown to “output” more information than was provided as “input” in the instructions for its construction, then the machine’s designer would have disproved Descartes’s dictum.

Without going too far into the technical details of information theory, the basic idea is that the quantity of information in a message is the measure of the reduction in uncertainty that results when the message is received. The technical definition differs significantly from the common-sense understanding of “information” insofar as the information contained in a message has nothing to do with the contents of the message itself, but only with the variety in the other messages from which it was selected, and so “information” is really a property of a system of communication rather than of any particular message within it. The reduction in uncertainty upon receiving a message thus depends on the probability of receiving the message, and also on the size of the set of possible messages to which it belongs.⁸ As the number of possible messages increases, either the number of different signals or the length of a message (composed of a sequence of signals) must also increase in order to make each message distinct from the others. In the binary encoding of computers, there are only two signals (or symbols), 0 and 1, and thus the length of the sequence needed to encode a message must increase as the number of possible messages increases in order for each message to be represented by a unique sequence.

Ashby used the theory of information to measure “design” by arguing that the choices made in a design are like the messages sent over a communication channel. That is, the significance of a choice is measured against the number of alternatives from which it must be selected. As he states it (Ashby 1952b, pp. 45–47):

How are we to obtain an objective and consistent measure of the “amount of design” put into, or shown by, a machine? Abstractly, “designing” a machine means giving selected numerical values to the available parameters. How long shall the lever be? where shall its fulcrum be placed? how many teeth shall the cog have? what value shall be given to the electrical resistance? what composition shall the alloy have? and so on. *Clearly, the amount of design must be related in some way to the number of decisions made and also to the fineness of the discrimination made in the selection* [emphasis added]. . . .

To apply the measure to a designed machine, we regard the machine as something specified by a designer and produced, as output, from a workshop. We must therefore consider not only the particular machine but the *ensemble* of machines from which the final model has been selected [original emphasis].

If one quantifies the information contained in a design as the choices made from among the possible alternatives, then one can make a similar move to quantify the information exhibited by the machine’s performance. The information displayed by the machine is the number of functionally distinct states it can exhibit—Ashby’s example is of a network consisting of a number of switches, the configuration of which determines different connectivities or states of the network. The design of the network is an assignment of values to the switches from among all the possible assignments. In this case, the network can only display as many states as the switches allow different configurations; some of the distinct assignments may be functionally equivalent and thus the machine may display *less* information than is contained in its design. But how, then, is it possible for a machine to display *more* information than is contained in its design?

The demonstration of this possibility draws close to the arguments about “design” during the rise of evolutionary theory in the nineteenth century. So close, in fact, that Ashby (1952b, p. 50) followed Norbert Wiener (1948) in calling instances of such systems “Darwinian Machinery”:

The question might seem settled, were it not for the fact, known to every biologist, that Descartes’ dictum was proved false over ninety years ago by Darwin. He showed that quite a simple rule, acting over a great length of time, could produce design and adaptation far more complex than the rule that had generated it. The status of his proof was uncertain for some time, but the work of the last thirty years, especially that of the geneticists, has shown beyond all reasonable doubt the sufficiency of natural selection. We face therefore something of a paradox. There can be no escape by denying the great complexity of living organisms. Neither Descartes nor Kant would have attempted this, for they appealed to just this richness of design as evidence for their arguments. Information theory, too, confirms this richness. Thus, suppose we try to measure the amount of design involved in the construction of a bird that can fly a hundred miles without resting. As a machine, it must have a very large number

of parameters adjusted. How many cannot be stated accurately, but it is of the same order as the number of all facts of avian anatomy, histology, and biochemistry. Unquestionably, therefore, evolution by natural selection produces great richness of design.

In evolution, there is an increasing amount of information displayed by the machine, despite the fact that the design is both simple and, in a sense, unchanging. Ashby (1952b) goes so far as to suggest that the design for a bird might be as simple as “Take a planet with some carbon and oxygen; irradiate it with sunshine and cosmic rays; and leave it alone for a few hundred million years” (p. 52). But the mechanism responsible for evolution is difficult to directly observe in action, and it does not appear to apply straightforwardly to a chess-playing machine.

If evolution is able to produce systems that exhibit more information than is contained in their design, and information cannot be spontaneously generated, where did this extra information come from? Obviously, this information must come in the form of an input of messages unforeseen by the designer (Ashby 1952b, p. 51):

The law that information cannot be created is not violated by evolution, for the evolving system receives an endless stream of information in the form of mutations. Whatever their origin, whether in cosmic rays or thermal noise, the fact that each gene may, during each second change unpredictably to some other form makes each gene a typical information source. The information received each second by the whole gene-pattern, or by the species, is then simply the sum of the separate contributions. The evolving system thus has *two* sources of information, that implied in the specifications of the rules of natural selection and that implied by the inpouring stream of mutations.

This philosophical problem was, of course, the same one which fueled much of the controversy over Darwin’s theory in the nineteenth century—whether the exquisite subtleties of living creatures could possibly be produced by brute natural processes or whether they necessarily required a supernatural “Designer.” What Darwin had so carefully detailed in *On the Origin of Species by Means of Natural Selection* (1859) was how natural evolutionary processes could lead to speciation—the divergence in forms of two distinct species who share a common ancestry; the branching of the tree of common descent. Assuming that the design of a species did not change in virtue of continuous divine intervention, the demonstration that species did change over time, and to such an extent as to result in new species, implied that natural evolutionary processes, in the absence of a designer, might have given rise to all biological forms. The basic process of natural selection choosing among the variations of form is argued to move species

toward those forms best able to survive and reproduce. Ashby simply placed a special emphasis on a portion of Darwin's theory by indicating how spontaneous variations in form provide an additional source of information apart from any determinate design.

In biological systems, the random variations of mutation supply alternative possibilities unforeseen by any designer, and thus the organism can evolve capacities beyond its own design. Similarly, Ashby (1952b) would argue, by adding a random number generator, Geiger counter, or other source of random noise to a system, we introduce the possibility of behaviors unforeseen in its "design" (p. 51):

It is now clear that the paradox arose simply because the words "cause" or "designer," in relation to a system, can be used in two senses. If they are used comprehensively, to mean "everything that contributes to the determination of the system," then Shannon and Descartes can agree that "a noiseless transducer or determinate machine can emit only such information as is supplied to it." This formulation will include the process of evolution if the "cause" is understood to include not only the rules of natural selection but also the mutations, *specified in every detail*. If, on the other hand, by "cause" or "designer" we mean something more restricted—a human designer, say—so that the designer is only a part of the total determination, then the dictum is no longer true.

With the paradox thus resolved, Ashby had demonstrated the possibility that a mechanical chess player *could* outplay its design(er). Further, he had identified the key to achieving this possibility, the flow of random information coming into the system. What remained to be shown was how this information could be made useful. A random move generator might contain the "supermoves" of chess, but how would a mechanical chess player be able to distinguish these moves from the rest? The answer to this question required developing a new conception of intelligence suitable to the mechanistic theory of mind.

Amplifying Intelligence

Once the analogy between adaptation and equilibrium was firmly set in Ashby's philosophy as the basis for a mechanistic theory of mind, he extended the analogy freely by describing mental processes using the terminology once reserved for describing machines such as steam engines and electronic devices: the engineer's language of "power," and "energy." One of his central themes in this respect was the application of the process of "amplification" to mental concepts such as intelligence. This extended analogy was not merely a rhetorical turn of phrase, but carried implications within his theoretical framework. Ashby thus turned his attention to devel-

oping a more rigorous definition of intelligence, and to demonstrating the significance of the mechanical-chess-player argument by showing how its results could be applied to practical problems. This line of thought culminated in his contribution to the first collected volume of work in the newly emerging subfields of computer science, artificial intelligence, and automata theory: Claude Shannon and John McCarthy's *Automata Studies*, published in 1956. The paper bore the intriguing title "Design for an Intelligence-Amplifier" and appeared in the final section of that volume, entitled "Synthesis of Automata." We will now examine that paper (Ashby 1956a) in detail and place its ideas in perspective with Ashby's overall philosophy.

Demonstrating that it was possible for a mechanical chess player to outplay its designer might be philosophically interesting, but showing that this discovery had practical significance would take more than arguments of metaphysical possibility. For this purpose, Ashby further extended his conception of the mechanisms of thought to problems of general interest, which took the form of a device that could "amplify" human intelligence. The continued reliance upon the analogy between thought and mechanical physics in his conception was made clear in the introduction to the paper (p. 215):

For over a century Man has been able to use, for his own advantage, physical powers that far transcend those produced by his own muscles. Is it impossible that he should develop machines with "synthetic" intellectual powers that will equally surpass those of his own brain? I hope to show that recent developments have made such machines possible—possible in the sense that their building can start today. Let us then consider the question of building a mechanistic system for the solution of problems that are beyond the human intellect. I hope to show that such a construction is by no means impossible, even though the constructors are themselves quite averagely human. There is certainly no lack of difficult problems awaiting solution. Mathematics provides plenty, and so does almost every branch of science. It is perhaps in the social and economic world that such problems occur most noticeably, both in regard to their complexity and to the great issues which depend on them. Success in solving these problems is a matter of some urgency. We have built a civilization beyond our understanding and we are finding that it is getting out of hand. Faced with such problems, what are we to do?

Rather than hope that individuals of extraordinary intelligence will step forward and solve such problems—a statistically unlikely eventuality—Ashby suggested that we ought to design machines that would amplify the intellectual powers of average humans. In the absence of careful definitions and criteria, such devices might sound quite fanciful. But with his usual

flare for mathematical rigor, Ashby provided those definitions and criteria and thereby also provided further illumination of his mechanistic philosophy of mind.

In resolving the problem of the mechanical chess player, Ashby had shown that a machine could output more information than was input through its design, by making use of other, random, information. This was a kind of amplification—information amplification—like the amplification of power that utilizes an input of power plus a source of free energy to output much more power than was originally supplied (p. 218):

[L]et us remember that the engineers of the middle ages, familiar with the principles of the lever and cog and pulley, must often have said that as no machine, worked by a man, could put out more work than he put in, therefore no machine could ever amplify a man's power. Yet today we see one man keeping all the wheels in a factory turning by shoveling coal into a furnace. It is instructive to notice just how it is that today's stoker defeats the mediaeval engineer's dictum, while being still subject to the law of the conservation of energy. A little thought shows that the process occurs in two stages. In Stage One the stoker lifts the coal into the furnace; and over this stage energy is conserved strictly. The arrival of the coal in the furnace is then the beginning of Stage Two, in which again energy is conserved, as the burning of the coal leads to the generation of steam and ultimately to the turning of the factory's wheels. By making the whole process, from stoker's muscles to factory wheel, take place in two stages, involving two lots of energy whose sizes can vary with some independence, the modern engineer can obtain an overall amplification.

In the mechanical chess player, as well as in evolution, information from the design, or problem specification, can be amplified in the same way that the strength of a stoker is amplified by a pile of coal and a steam engine, by the addition of free energy or random information. But the availability of bare information is not in itself intelligence, any more than free energy is work—these resources must be directed toward a task or goal.

What then is a suitable criterion for intelligent behavior? By starting from a definition of information that considered only its technical implications, a definition that leaves information independent of any analysis of it, Ashby was able to take account of analysis and judgment in his definition of intelligence. According to Ashby, intelligence implies a selection: *intelligence is the power of appropriate selection*. To see what this means, consider his example (p. 217):

It has often been remarked that any random sequence, if long enough, will contain *all* the answers. Nothing prevents a child from doodling " $\cos^2 x + \sin^2 x = 1$," or a dancing mote in the sunlight from emitting the same message in Morse or a similar code. Let us be more definite. If each of the above thirteen symbols might have been

any one of fifty letters and elementary signs, then as 50^{13} is approximately 2^{73} , the equation can be given in coded form by 73 binary symbols. Now consider a cubic centimeter of air as a turmoil of colliding molecules. A particular molecule's turnings after collision, sometimes to the left and sometimes to the right, will provide a series of binary symbols, each 73 of which, on some given code, either will or will not represent the equation. A simple calculation from the known facts shows that the molecules in every cubic centimeter of air are emitting this sequence correctly over a hundred thousand times a second. The objection that "such things don't happen" cannot stand. Doodling, then, or any other random activity, is capable of producing all that is required. What spoils the child's claim to be a mathematician is that he will doodle, with equal readiness, such forms as " $\cos^2 x + \sin^2 x = 2$ " or "ci)xi-nx1" or any other variation. After the child has had some mathematical experience he will stop producing these other variations. *He becomes not more, but less productive: he becomes selective.* [emphasis added]

In order to be intelligent, a mechanism must exhibit discipline in its behavior. Thus, given an ample source of random information, the efforts toward designing an intelligence amplifier ought to focus on the mechanisms of appropriate selection by which the device can choose which among the many possibilities is the desired answer. This definition constitutes a kind of inversion of the common formulation of machine intelligence understood as the ability to produce correct responses by design; intelligence is now understood as a combination of the abilities to *produce a great many meaningless alternatives*, and to *eliminate by appropriate selection the incorrect choices among those*—a two-stage process.

Exactly how to construct a mechanism to make appropriate selections thus becomes the design problem for building an intelligence amplifier. The design of an intelligent selector involves two major parts. The first is to establish criteria of selection that can be utilized by the machine, sufficient for it to know when it has arrived at an acceptable solution to the given problem. The second part involves coupling the selector to a source of chaotic information which it can search through in order to find an acceptable solution (p. 223):

Consider the engineer who has, say, some ore at the foot of a mine-shaft and who wants it brought to the surface. The power required is more than he can supply personally. What he does is to take some system that is going to change, by the laws of nature, from low entropy to high, and he couples this system to his ore, perhaps through pistons and ropes, so that "low entropy" is coupled to "ore down" and "high entropy" to "ore up." He then lets the whole system go, confident that as the entropy goes from low to high so will it change the ore's position from down to up. Abstractly... he has a process that is going, by the laws of nature, to pass from state H_1 to state H_2 . He wants C_1 to change to C_2 . So he couples H_1 to C_1 and H_2 to

C_2 . Then the system, in changing from H_1 to H_2 , will change C_1 to C_2 , which is what he wants. The arrangement is clearly both necessary and sufficient. The method of getting the problem-solver to solve the set problem can now be seen to be of essentially the same form. The job to be done is the bringing of $X \dots$ to a certain condition or "solution" \sim . What the intelligence engineer does first is build a system, X and S , that has the tendency, by the laws of nature, to go to a state of equilibrium. He arranges the coupling between them so that "not at equilibrium" is coupled to not \sim , and "at equilibrium" to \sim . He then lets the system go, confident that as the passage of time takes the whole to an equilibrium, so will the conditions in X have to change from not \sim to \sim . He does not make the conditions in X change by his own efforts, but allows the basic drive of nature to do the work. This is the fundamental principle of our intelligence-amplifier. Its driving power is the tendency for entropy to increase, where "entropy" is used, not as understood in heat-engines, but as understood in stochastic processes.

In yet another inversion of traditional thought, Ashby has demonstrated how the natural processes of entropy in nature, the relentless destruction of organization, can be used as the fuel for the amplification of intelligence beyond the capabilities of the naked human mind.

The key to intelligence thus lies in selectivity, for it is the power of appropriate selection that is able to recognize the desired messages from among the chaos of random information. But how does one achieve this in a machine? Consider, as Ashby does, a machine to solve difficult social and economic problems. As designers, we make *our* selection as to what we want, say (p. 219):

An organisation that will be stable at the conditions:

Unemployed	<100,000 persons
Crimes of violence	<10 per week
Minimal income per family	>£500 per annum

Taking these desiderata as the machine's goal, it is the task of the machine to sift through an enormous number of possible economic configurations, and select one that meets these conditions. Part of the design of that machine involves specifying the representation of the economic system, and thus the set of things from which the selection must take place. Apart from this, Ashby has little to say about this design process—a topic with which much of the work in artificial intelligence has since been concerned. But herein lies another essential point, for it raises again the question of information. This is to say that in determining the class of things from which a selection is to be made one also specifies the amount of information that the answer will require. Since the measure of the information contained in a message is the reduction in uncertainty resulting from the message

being received, by determining the size of the set of possible messages—answers—the designer has put a number on the amount of information needed to solve the problem.

In later writings, Ashby returned to this problem and gave it a proper formalization using information theory. That formulation involved seeing the process of selection not as an instance of the perfect transmission of information but as a form of communication over a noisy channel. In so doing, he saw a deep and interesting connection between Shannon's 10th Theorem (1948) and his own Law of Requisite Variety (Ashby 1956b, p. 202).⁹ The formulation involves equating the entropic source of random information with a noisy channel, and selection with the problem of determining which messages are correct and which are not. In order for someone on the receiving end of a noisy channel to determine the correctness of a message, they must receive an additional source of information, a kind of feedback regarding the correctness of the messages received. This information comes through an error-correcting channel. Shannon's 10th Theorem provides a measure of the capacity of the channel necessary to achieve error-free transmission over a noisy channel (within a certain degree of accuracy). Ashby argued that in order to make a correct selection in a decision process, a system must receive information from the environment and that the measure of this information is equivalent to the required capacity for an error-correcting channel (Ashby 1960, p. 746).

To see what this means, consider the case in which the number of possible economic configurations our problem solver must select from is 1,000,001, and there is only one correct solution. Suppose that it is possible to eliminate whole classes or subsets of this set as inappropriate. A message on the error-correcting channel transmits this information by indicating a single subset that the correct answer cannot be a part of. Let us say that each subset in our problem contains exactly 1,000 unique economic configurations (in most real problems the size of each subset is different and many subsets overlap and share members, but we shall ignore these difficulties). In this case every message eliminates a thousand possibilities, leaving the selector with 999,001 possibilities after the first message, and then with 998,001 after the second message, and so on. At this rate, it will take at least 1,000 messages to achieve complete certainty that the selector will have the right answer, but fewer if we do not require 100 percent certainty. At each step it has made some progress as the probability of correctness for each of the answers still in the set of possibilities goes up after each piece of information is received. But when it comes to choosing from among the elements remaining, the selector has no more information available for

deciding whether any one of the remaining elements is “better” or “worse” than any of the others—it can only pick one at random. If the selector had more information and were thus able to make a selection among the remaining elements, it would do so until it was again left with a set of elements where each was no more likely to be correct than any other.

This led Ashby to the conclusion that all forms of intelligence depend necessarily on receiving information in order to achieve any appropriate selection that they make. And the greater the set of possibilities and complexity of the partitioning of alternatives, the more information will be required for the selection to be appropriate. No intelligence is able to create a brilliant idea from nothing; genius of this sort is merely a myth (Ashby 1961, p. 279):

Is there, then, no such thing as “real” intelligence? What I am saying is that if by “real” one means the intelligence that can perform great feats of appropriate selection without prior reception and processing of the equivalent quantity of information; then such “real” intelligence does not exist. It is a myth. It has come into existence in the same way that the idea of “real” magic comes to a child who sees conjuring tricks.

When humans appear to achieve remarkable performances of “genius,” it is only because they had previously processed the required amount of information. Ashby argues that were it possible for such selections to occur in the absence of the required information processing, it would be like the case of a student who provided answers to exam questions before they were given—it would upset the causal order (Ashby 1960, p. 746).

When considering whether a machine such as a computer is capable of selective—that is, intelligent—performances at the level of skill of the human mind, he warns that we must carefully note how much information has been processed by each system (Ashby 1961, pp. 277–278):

It may perhaps be of interest to turn aside for the moment to glance at the reasons that may have led us to misunderstand the nature of human intelligence and cleverness. The point seems to be, as we can now see with the clearer quantitative grasp that we have today, that we tended grossly to mis-estimate the quantities of information that were used by computers and by people. When we program a computer, we have to write down every detail of the supplied information, and we are acutely aware of the quantity of information that must be made available to it. As a result, we tend to think that the quantity of information is extremely large; in fact, on any comparable scale of measurement it is quite small. The human mathematician, however, who solves a problem in three-dimensional geometry for instance, may do it very quickly and easily, and he may think that the amount of information that he

has used is quite small. In fact, it is very large; and the measure of its largeness is precisely the amount of programming that would have to go into the computer in order to enable the computer to carry through the same process and to arrive at the same answer. The point is, of course, that when it comes to things like three-dimensional geometry, the human being has within himself an enormous quantity of information obtained by a form of preprogramming. Before he picked up his pencil, he already had behind him many years of childhood, in which he moved his arms and legs in three-dimensional space until he had learned a great deal about the intricacies of its metric. Then he spent years at school, learning formal Euclidian methods. He has done carpentry, and has learned how to make simple boxes and three-dimensional furniture. And behind him is five billion years of evolutionary molding all occurring in three-dimensional space; because it induced the survival of those organisms with an organisation suited to three-dimensional space rather than to any other of the metrics that the cerebral cortex could hold. . . . What I am saying is that if the measure is applied to both on a similar basis it will be found that each, computer and living brain, can achieve appropriate selection precisely so far as it is allowed to by the quantity of information that it has received and processed.

Once formulated in this way, we can recognize certain connections to aspects of Ashby's philosophy discussed earlier in this chapter. Most obvious is the significance of evolutionary adaptation as a source of information. On the one hand, there are the countless random trials and errors of that history—the raw information of random variation. But there is also the resultant information of selective adaptation: what was won from those trials and errors was a better organization for dealing with the environment. For the mathematician, that organization is already a part of him. As a model of the evolutionary history of his species, and of his own life experiences, he stands as an *archive* of that information—it is embodied in his cerebral organization. For the computer, the programmer stands as a designer who must make each of those decisions necessary for the mathematician's performance and express them in a computer program. It would be more desirable for the machine to learn those things itself, but this merely means that the information comes from a different source, not that it is spontaneously created by the machine.

With an account of the process of appropriate selection that was sufficient for quantitative measurement, Ashby had completed his general outline of a mechanistic philosophy of mind. It formed the basis, he believed, for an objective scientific study of intelligence. It provided in its formal rigor a means for experimentation and observations capable of resolving theoretical disputes about the mind. It also provided a basis for the synthesis of mechanical devices capable of achieving adaptive and intelligent

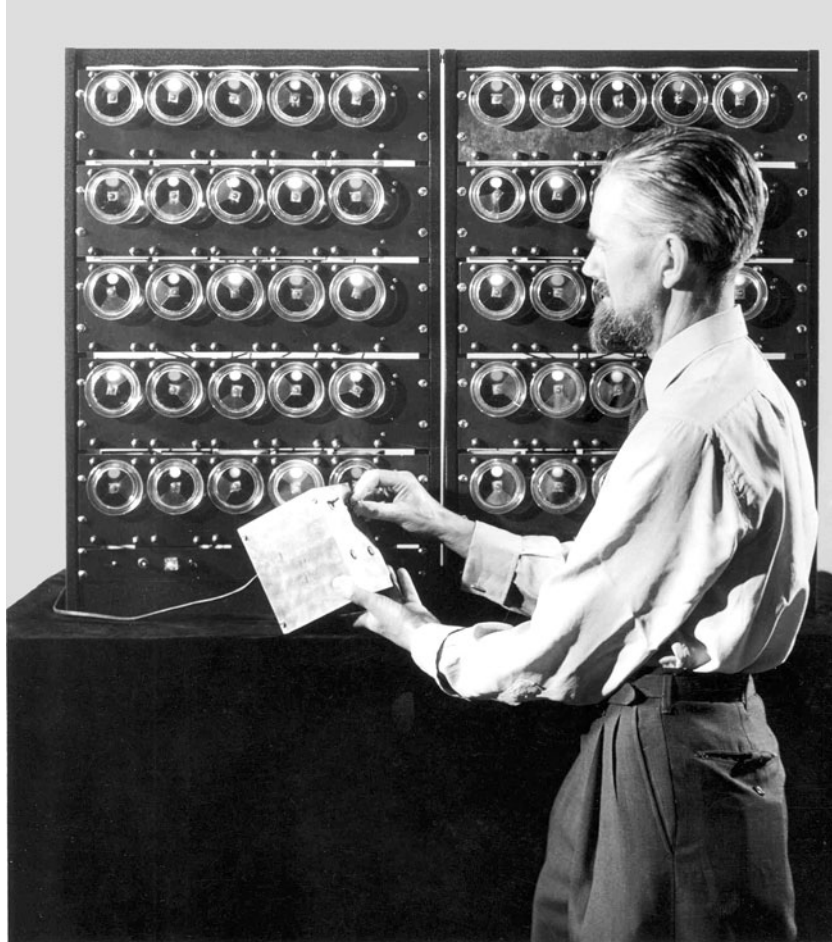


Figure 7.2

Ashby at the Biological Computer Laboratory (BCL), University of Illinois, with his "Grandfather Clock" and "Non-Trivial Machine." Used with permission of Murray Babcock's widow, Velva Babcock.

performances; the Homeostat was only one of the devices capable of such performances that Ashby constructed. His theoretical framework brought together physical, biological, and psychological theory in a novel and powerful form, one that he would credit Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow (1943) and G. Sommerhoff (1950) for having independently discovered in their own work (Ashby 1952c). He would also agree that his conception of “adaptation and equilibrium” was equivalent to Sommerhoff’s “directive correlation” and Rosenblueth, Wiener, and Bigelow’s conception of “negative feedback”—the central concept of cybernetics. But Ashby also extended this idea to the more subtle aspects of intelligence: How could human intelligence be extended by machines? And what were the mechanics of decision-making processes?

Conclusion

Ashby’s mechanistic philosophy of mind bears many superficial similarities to the more popular formulations of the idea that “machines can think,” in particular the formulation provided by the “Turing test.” Now that we have examined Ashby’s philosophy in its details, however, it is instructive to note the subtle differences. The demonstration of the fundamental equivalence of adaptation and equilibrium was the core of Ashby’s conception of the mind as a mechanism. Although Alan Turing demonstrated (1936) that any formally describable process could be performed by a computer, he recognized that this was not itself sufficient to show that a computer could think, since thinking might not be a formally describable process. Moreover, it did not come close to explaining *how* a computer could think. Ashby had set himself a different task than Turing: to understand how the behaviors and performances of living organisms in general, and thinking brains in particular, could be composed of mechanisms at all, and what those mechanisms were.

Consider Turing’s (1950) “imitation game” for deciding whether or not a machine could be intelligent. In the first sections of that paper, he completely avoids attempting to define “machine” or “intelligence.” Instead, he insists with little argument that the machine must be a digital computer, and proceeds to substitute his imitation game for a formal definition of intelligence. While we might agree with Turing that appealing to a commonsense understanding of “intelligence” would amount to letting the truth of the statement “intelligent machines can be made” depend upon the popular acceptance of the statement, his own imitation game doesn’t go much further than this. In Turing’s test for intelligence, he pits a digital

computer against a real human being in a game where the winning objective for all contestants is to convince human judges that they are the humans and not the computer. The computer is considered “intelligent” if it is able to convince more than 50 percent of the judges that it is the human. Turing sets out some rules, to ensure that digital computers can play on an even field, which require that all interactions between the judges and the contestants take place over a telegraph wire, which limits the intelligent performances to the output of strings of symbols. Much has been written about this “test” for machine intelligence, and it is certainly the most popular formulation of the problem, but it seems profoundly lacking when compared to Ashby’s definition of machine intelligence (and even the other ideas offered by Turing).

First, the fact that the “common usage” of the term “intelligence” is insufficient for judging computers does not mean that a precise formal definition cannot be provided—indeed, this is just what Ashby believed he had done. Second, the restriction of the meaning of “machines” to “digital computers” seems unnecessary. The Homeostat, for one, is an analogue computer that seems quite capable of demonstrating an intelligent capacity. Moreover, it does so not by virtue of carrying out particular calculations but of being a certain kind of information-processing system, one that is goal-seeking and adaptive. More significant, by leaving the meaning of intelligence up to a population of judges with indeterminate criteria, Turing’s test fails to offer any instruction as to how such a computer should be constructed, or what its specific intellectual capacities might be—it is a way to dodge the issue of what intelligence is altogether.

In the process of developing his mechanistic philosophy, Ashby managed to perform some inversions of intuitions that are still commonly held. The first of these inversions was the “generative power of breakdown.” The idea that creation requires impermanence, that destruction precedes construction, or that from chaos comes order is a recurring paradox of metaphysical thought, at least as ancient as Greek. In another form, it reappears in Ashby’s work as a system’s need for a source of random information in order to achieve a better performance than it was previously capable of. And it appears again when entropy is used as the fuel for driving the intelligence-amplifier to superhuman performances of appropriate selection. The intelligence-amplifier also inverts the notion that originality and productivity are essential aspects of intelligence. These are aspects of the random information fed to a selector, but it is the *power of appropriate selection* that reduces productivity and originality in a highly disciplined process which gives only the desired result.

To the end of his career Ashby remained concerned with the specific requirements for building machines that exhibited brainlike behavior. In part, this was motivated by his desire to understand the brain and its processes, and in part it was to build machines capable of aiding the human intellect. Although his designs for an intelligence-amplifier may still sound fanciful, his belief that such machines could be usefully brought to bear on real economic and social problems was not (Ashby 1948, pp. 382–83):

The construction of a machine which would react successfully to situations more complex than can be handled at present by the human brain would transform many of our present difficulties and perplexities. Such a machine might be used, in the distant future, not merely to get a quick answer to a difficult question, but to explore regions of intellectual subtlety and complexity at present beyond the human powers. The world's political and economic problems, for instance, seem sometimes to involve complexities beyond even the experts. Such a machine might perhaps be fed with vast tables of statistics, with volumes of scientific facts and other data, so that after a time it might emit as output a vast and intricate set of instructions, rather meaningless to those who had to obey them, yet leading, in fact, to a gradual resolving of the political and economic difficulties by its understanding and use of principles and natural laws which are to us yet obscure. The advantages of such a machine are obvious. But what of its disadvantages?

His aim was thus not merely to understand the brain, and simulate its properties, but also to understand those properties in such a way that they could be usefully employed to resolve difficult intellectual problems.

Even while he held out a hopeful vision of a future in which intelligent machines could resolve problems of great human concern and consequence, he was not without his fears of what the actual results might be (Ashby 1948). An intelligent machine by his definition was, after all, a machine that succeeded in achieving its *own purposes*, regardless of the resistance it encountered (p. 383):

But perhaps the most serious danger in such a machine will be its selfishness. Whatever the problem, it will judge the appropriateness of an action by how the feedback affects itself: not by the way the action benefits us. It is easy to deal with this when the machine's behavior is simple enough for us to be able to understand it. The slave-brain will give no trouble. But what of the homeostat-type, which is to develop beyond us? In the early stages of its training we shall doubtless condition it heavily to act so as to benefit ourselves as much as possible. But if the machine really develops its own powers, it is bound sooner or later to recover from this. If now such a machine is used for large-scale social planning and coordination, we must not be surprised if we find after a time that the streams of orders, plans and directives issuing from it begin to pay increased attention to securing its own welfare. Matters like the

supplies of power and the prices of valves affect it directly and it cannot, if it is a sensible machine, ignore them. Later, when our world-community is entirely dependent on the machine for advanced social and economic planning, we would accept only as reasonable its suggestion that it should be buried deeply for safety. We would be persuaded of the desirability of locking the switches for its power supplies permanently in the "on" position. We could hardly object if we find that more and more of the national budget (planned by the machine) is being devoted to ever-increasing developments of the planning-machine. In the spate of plans and directives issuing from it we might hardly notice that the automatic valve-making factories are to be moved so as to deliver directly into its own automatic valve-replacing gear; we might hardly notice that its new power supplies are to come directly from its own automatic atomic piles; we might not realise that it had already decided that its human attendants were no longer necessary. How will it end? I suggest that the simplest way to find out is to make the thing and see.

This vision of the evolution of machines is sobering and sounds like the stuff of science fiction. In fact, however, it is more reserved than many of the claims made in the fields of artificial life and Artificial Intelligence in six decades since it was written. More to the point, when viewed in perspective with Ashby's overall philosophy it provides a means for thinking about the processes of social and economic organization and planning with a particular emphasis on the flow of information in those processes; though Ashby did not pursue this idea, it would seem to warrant further study.

There are many subtleties, implications, and extensions of Ashby's mechanistic philosophy that we have not covered. There are also many aspects of his intellectual career and contributions that we have skipped over or touched on only briefly. Our aim, however, was to come to a much clearer view of Ashby's overall philosophy, and of the interconnections and dependencies between its elements, so as to gain a greater appreciation for what is contained in Ashby's idea of "mechanical intelligence."

Notes

1. Both books were translated into several languages: *Design For a Brain* was published in Russian (1958), Spanish (1959), and Japanese (1963); *An Introduction to Cybernetics* was published in Russian (1957), French (1957), Spanish (1958), Czech (1959), Polish (1959), Hungarian (1959), German (1965), Bulgarian (1966), and Italian (1966).
2. Though it is implicit in much of AI, this approach is most explicit in the current field of biorobotics (see Webb and Consi 2001), and was also central in the develop-

ment of the fields of bionics and self-organizing systems in the 1960s (see Asaro 2007; for more on the synthetic method in the work of Ashby and a fellow cybernetician, W. Grey Walter, see Asaro 2006).

3. It is interesting to note that advantage 2 in this summary presages A. Rosenblueth, Norbert Wiener, and Julian Bigelow's (1943) "Behavior, Purpose, and Teleology" by three years. Ashby also bases his arguments on an elaboration of the concept of a "functional circuit," emphasizing the stable type, which parallels Rosenblueth, Wiener, and Bigelow's concept of feedback mechanisms, and negative feedback in particular, as explaining purposive or goal-seeking behavior. Another researcher, G. Sommerhoff (1950), a physicist attempting to account for biological organisms as physical systems, would come to essentially the same concepts a few years later. In his review of Sommerhoff's *Analytical Biology* Ashby (1952e) himself concludes, "It shows convincingly that the rather subtle concept of 'adaptation' can be given a definition that does full justice to the element of 'purpose,' while achieving a degree of precision and objectivity hitherto obtainable only in physics. As three sets of workers have now arrived independently at a definition from which the individuals differ only in details, we may reasonably deduce that the concept of 'adaptation' *can* be so re-formulated, and that its formulation in the language of physics is now available" (p. 409).

4. See Ashby (1947), "The Nervous System as Physical Machine: With Special Reference to the Origin of Adaptive Behavior," for more on learning and adaptation in the kitten.

5. It is interesting to note as an aside that, despite his relentless use of "stability" and later coining of the terms "ultrastability," "poly-stable" and "multi-stable," he does not use the word at all in his second paper on the mechanisms of adaptation, "The Physical Origin of Adaptation by Trial and Error" (1945; submitted 1943). There he uses the term "normal" in the place of "stability." This was perhaps due to a difference in audiences since this paper was addressed to psychologists.

6. Bigelow was a colleague of Norbert Wiener's at MIT, and was a coauthor of "Behavior, Purpose, and Teleology" (1943), which marks the beginning of cybernetics. He was the electrical engineer who built Wiener's "anti-aircraft predictor." In 1946 he had become the chief engineer of John von Neumann's machine at Princeton's Institute for Advanced Study, one of the first stored-program electronic computers.

7. Descartes's dictum can be found in the *Meditations*, and is a premise in his argument for the existence of God. The other premise is that "I find upon reflection that I have an idea of God, as an infinitely perfect being," from which Descartes concludes that he could not have been the cause of this idea, since it contains more perfection than he does, and thus there must exist an infinitely perfect God which is the real cause of his idea of an infinitely perfect God. He goes on to argue that the same God endowed him with reliable perception of the world.

8. Shannon's (1948) equation for the quantity of information is: $-\sum_j p_j \log p_j$, where p_j is the probability of receiving message j . By summing over all the messages, we obtain a measure of the current uncertainty, and thus of how much uncertainty will be removed when we actually receive a message and become certain. Thus the uncertainty is a measure of the system of communication and is not really a property of the message; alternatively we could say that the information content is the same for equiprobable messages in the set.

9. Shannon's 10th Theorem (1948, p. 68) states: "If the correction channel has a capacity equal to $H_y(x)$ it is possible to so encode the correction data as to send it over this channel and correct all but an arbitrarily small fraction ε of the errors. This is not possible if the channel capacity is less than $H_y(x)$." Here $H_y(x)$ is the conditional entropy of the input (x) when the output (y) is known.

Ashby's Law of Requisite Variety states that any system that is to control the ultimate outcome of any interaction in which another system also exerts some control must have at least as much variety in its set of alternative moves as the other system if it is to possibly succeed (Ashby 1956b, p. 206).

References

- Asaro, Peter. 2006. "Working Models and the Synthetic Method: Electronic Brains as Mediators Between Neurons and Behavior." *Science Studies* 19(1): 12–34.
- . 2007. "Heinz von Foerster, the Bio-Computing Movements of the 1960s, and the Machines of the BCL." In *A Disrupted Revolution? Heinz von Foerster and the Biological Computer Laboratory, 1958–1974*, edited by Albert Müller and Karl H. Müller. Vienna: Edition Echoraum.
- Ashby, W. Ross. 1940. "Adaptiveness and Equilibrium." *Journal of Mental Science* 86: 478–83.
- . 1945. "The Physical Origin of Adaptation by Trial and Error." *Journal of General Psychology* 32: 13–25.
- . 1947. "The Nervous System as Physical Machine: With Special Reference to the Origin of Adaptive Behavior." *Mind* 56, no. 1: pp. 44–59.
- . 1948. "Design for a Brain." *Electronic Engineering* 20: 379–83.
- . 1951. "Statistical Machinery." *Thales* 7: 1–8.
- . 1952a. "Homeostasis." In *Cybernetics: Transactions of the Ninth Conference*, edited by H. von Foerster. New York: Josiah Macy Jr. Foundation (March), 73–108.
- . 1952b. "Can a Mechanical Chess-player Outplay Its Designer?" *British Journal for the Philosophy of Science* 3, no. 9: 44–57.

- . 1952c. *Design for a Brain*. London: Chapman & Hall.
- . 1952d. "Mechanical Chess Player." In *Cybernetics: Transactions of the Ninth Conference*, edited by H. von Foerster. New York: Josiah Macy Jr. Foundation (March), 151–54.
- . 1952e. "Review of *Analytical Biology*, by G. Sommerhoff." *Journal of Mental Science* 98: 408–9.
- . 1954. "Review of *The Neurophysiological Basis of Mind*, by J. C. Eccles." *Journal of Mental Science* 100: 511.
- . 1956a. "Design for an Intelligence-Amplifier." In *Automata Studies*, edited by Claude E. Shannon and J. McCarthy. Princeton: Princeton University Press.
- . 1956b. *An Introduction to Cybernetics*. London: Chapman & Hall.
- . 1960. "Computers and Decision Making." *New Scientist* 7: 746.
- . 1961. "What Is an Intelligent Machine?" BCL technical report no. 7.1. Urbana: University of Illinois, Biological Computer Laboratory.
- . 1962. "Simulation of a Brain." In *Computer Applications in the Behavioral Sciences*, edited by H. Borko. New York: Plenum Press.
- . 1967. "Cybernetics of the Large System." In *Accomplishment Summary 1966/67*. BCL report no. 67.2. Urbana: University of Illinois, Biological Computer Laboratory.
- Cordeschi, Roberto. 2002. *The Discovery of the Artificial*. Dordrecht: Kluwer.
- Jennings, H. S. 1915. *Behavior of the Lower Organisms*. New York: Columbia University Press.
- Loeb, J. 1900. *Comparative Physiology of the Brain and Comparative Psychology*. New York: G. P. Putnam's Sons.
- Lorentz, H. A. 1927. *Theoretical Physics*. London: Macmillan.
- Miessner, B. F. 1916. *Radiodynamics: The Wireless Control of Torpedoes and Other Mechanisms*. New York: Van Nostrand.
- Rosenblueth, A., Norbert Wiener, and Julian Bigelow. 1943. "Behavior, Purpose, and Teleology." *Philosophy of Science* 10: 18.
- Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27: 379–423 and 623–56.
- Shannon, Claude E., and J. McCarthy, eds. 1956. *Automata Studies*. Princeton: Princeton University Press.
- Sommerhoff, G. 1950. *Analytical Biology*. London: Oxford University Press.

Turing, Alan M. 1936. "On Countable Numbers, with an Application to the *Entscheidungsproblem*." *Proceedings of the London Mathematics Society* (series 2) 42: 230–65.

———. 1950. "Computing Machinery and Intelligence." *Mind* 59: 433–60.

Webb, Barbara, and Thomas R. Consi (Eds.). 2001. *Biorobotics*. Cambridge, Mass.: MIT Press.

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