Herbert A. Simon

The Sciences of the Artificial

Third Edition

The Sciences of the Artificial

Third edition

Herbert A. Simon

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Sabon by Graphic Composition, Inc.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Simon, Herbert Alexander, 1916–
The sciences of the artificial / Herbert A. Simon.—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-19374-4 (alk. paper).—ISBN 0-262-69191-4 (pbk.: alk. paper)
1. Science—Philosophy. I. Title.
Q175.S564 1996
300.1'1—dc20 96-12633
CIP

10 9 8 7 6

1

Understanding the Natural and the Artificial Worlds

About three centuries after Newton we are thoroughly familiar with the concept of natural science—most unequivocally with physical and biological science. A natural science is a body of knowledge about some class of things—objects or phenomena—in the world: about the characteristics and properties that they have; about how they behave and interact with each other.

The central task of a natural science is to make the wonderful commonplace: to show that complexity, correctly viewed, is only a mask for simplicity; to find pattern hidden in apparent chaos. The early Dutch physicist Simon Stevin, showed by an elegant drawing (figure 1) that the law of the inclined plane follows in "self-evident fashion" from the impossibility of perpetual motion, for experience and reason tell us that the chain of balls in the figure would rotate neither to right nor to left but would remain at rest. (Since rotation changes nothing in the figure, if the chain moved at all, it would move perpetually.) Since the pendant part of the chain hangs symmetrically, we can snip it off without disturbing the equilibrium. But now the balls on the long side of the plane balance those on the shorter, steeper side, and their relative numbers are in inverse ratio to the sines of the angles at which the planes are inclined.

Stevin was so pleased with his construction that he incorporated it into a vignette, inscribing above it

Wonder, en is gheen wonder

that is to say: "Wonderful, but not incomprehensible."

This is the task of natural science: to show that the wonderful is not incomprehensible, to show how it can be comprehended—but not to

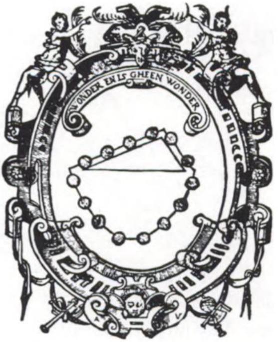

Figure 1
The vignette devised by Simon Stevin to illustrate his derivation of the law of the inclined plane

destroy wonder. For when we have explained the wonderful, unmasked the hidden pattern, a new wonder arises at how complexity was woven out of simplicity. The aesthetics of natural science and mathematics is at one with the aesthetics of music and painting—both inhere in the discovery of a partially concealed pattern.

The world we live in today is much more a man-made,[1] or artificial, world than it is a natural world. Almost every element in our environment shows evidence of human artifice. The temperature in which we spend most of our hours is kept artificially at 20 degrees Celsius; the humidity is added to or taken from the air we breathe; and the impurities we inhale are largely produced (and filtered) by man.

Moreover for most of us—the white-collared ones—the significant part of the environment consists mostly of strings of artifacts called "symbols" that we receive through eyes and ears in the form of written and spoken language and that we pour out into the environment—as I am now doing—by mouth or hand. The laws that govern these strings of

---

[1] I will occasionally use "man" as an androgynous noun, encompassing both sexes, and "he," "his," and "him" as androgynous pronouns including women and men equally in their scope.

symbols, the laws that govern the occasions on which we emit and receive them, the determinants of their content are all consequences of our collective artifice.

One may object that I exaggerate the artificiality of our world. Man must obey the law of gravity as surely as does a stone, and as a living organism man must depend for food, and in many other ways, on the world of biological phenomena. I shall plead guilty to overstatement, while protesting that the exaggeration is slight. To say that an astronaut, or even an airplane pilot, is obeying the law of gravity, hence is a perfectly natural phenomenon, is true, but its truth calls for some sophistication in what we mean by "obeying" a natural law. Aristotle did not think it natural for heavy things to rise or light ones to fall (*Physics,* Book IV); but presumably we have a deeper understanding of "natural" than he did.

So too we must be careful about equating "biological" with "natural." A forest may be a phenomenon of nature; a farm certainly is not. The very species upon which we depend for our food—our corn and our cattle—are artifacts of our ingenuity. A plowed field is no more part of nature than an asphalted street—and no less.

These examples set the terms of our problem, for those things we call artifacts are not apart from nature. They have no dispensation to ignore or violate natural law. At the same time they are adapted to human goals and purposes. They are what they are in order to satisfy our desire to fly or to eat well. As our aims change, so too do our artifacts—and vice versa.

If science is to encompass these objects and phenomena in which human purpose as well as natural law are embodied, it must have means for relating these two disparate components. The character of these means and their implications for certain areas of knowledge—economics, psychology, and design in particular—are the central concern of this book.

The Artificial

Natural science is knowledge about natural objects and phenomena. We ask whether there cannot also be "artificial" science—knowledge about artificial objects and phenomena. Unfortunately the term "artificial" has a pejorative air about it that we must dispel before we can proceed.

My dictionary defines "artificial" as, "Produced by art rather than by nature; not genuine or natural; affected; not pertaining to the essence of the matter." It proposes, as synonyms: affected, factitious, manufactured, pretended, sham, simulated, spurious, trumped up, unnatural. As antonyms, it lists: actual, genuine, honest, natural, real, truthful, unaffected. Our language seems to reflect man's deep distrust of his own products. I shall not try to assess the validity of that evaluation or explore its possible psychological roots. But you will have to understand me as using "artificial" in as neutral a sense as possible, as meaning man-made as opposed to natural.[2]

In some contexts we make a distinction between "artificial" and "synthetic." For example, a gem made of glass colored to resemble sapphire would be called artificial, while a man-made gem chemically indistinguishable from sapphire would be called synthetic. A similar distinction is often made between "artificial" and "synthetic" rubber. Thus some artificial things are imitations of things in nature, and the imitation may use either the same basic materials as those in the natural object or quite different materials.

As soon as we introduce "synthesis" as well as "artifice," we enter the realm of engineering. For "synthetic" is often used in the broader sense of "designed" or "composed." We speak of engineering as concerned with "synthesis," while science is concerned with "analysis." Synthetic or artificial objects—and more specifically prospective artificial objects having desired properties—are the central objective of engineering activity and skill. The engineer, and more generally the designer, is concerned with how things *ought* to be—how they ought to be in order to *attain goals,*

---

2. I shall disclaim responsibility for this particular choice of terms. The phrase "artificial intelligence," which led me to it, was coined, I think, right on the Charles River, at MIT. Our own research group at Rand and Carnegie Mellon University have preferred phrases like "complex information processing" and "simulation of cognitive processes." But then we run into new terminological difficulties, for the dictionary also says that "to simulate" means "to assume or have the mere appearance or form of, without the reality; imitate; counterfeit; pretend." At any rate, "artificial intelligence" seems to be here to stay, and it may prove easier to cleanse the phrase than to dispense with it. In time it will become sufficiently idiomatic that it will no longer be the target of cheap rhetoric.

and to *function.* Hence a science of the artificial will be closely akin to a science of engineering—but very different, as we shall see in my fifth chapter, from what goes currently by the name of "engineering science."

With goals and "oughts" we also introduce into the picture the dichotomy between normative and descriptive. Natural science has found a way to exclude the normative and to concern itself solely with how things are. Can or should we maintain this exclusion when we move from natural to artificial phenomena, from analysis to synthesis?[3]

We have now identified four indicia that distinguish the artificial from the natural; hence we can set the boundaries for sciences of the artificial:

1. Artificial things are synthesized (though not always or usually with full forethought) by human beings.
2. Artificial things may imitate appearances in natural things while lacking, in one or many respects, the reality of the latter.
3. Artificial things can be characterized in terms of functions, goals, adaptation.
4. Artificial things are often discussed, particularly when they are being designed, in terms of imperatives as well as descriptives.

The Environment as Mold

Let us look a little more closely at the functional or purposeful aspect of artificial things. Fulfillment of purpose or adaptation to a goal involves a relation among three terms: the purpose or goal, the character of the artifact, and the environment in which the artifact performs. When we think of a clock, for example, in terms of purpose we may use the child's definition: "a clock is to tell time." When we focus our attention on the clock itself, we may describe it in terms of arrangements of gears and the

---

[3] This issue will also be discussed at length in my fifth chapter. In order not to keep readers in suspense, I may say that I hold to the pristine empiricist's position of the irreducibility of "ought" to "is," as in chapter 3 of my *Administrative Behavior* (New York: Macmillan, 1976). This position is entirely consistent with treating natural or artificial goal-seeking systems as phenomena, without commitment to their goals. *Ibid.*, appendix. See also the well-known paper by A. Rosenbluth, N. Wiener, and J. Bigelow, "Behavior, Purpose, and Teleology," *Philosophy of Science, 10* (1943):18–24.

Analogous to the role played by natural selection in evolutionary biology is the role played by rationality in the sciences of human behavior. If we know of a business organization only that it is a profit-maximizing system, we can often predict how its behavior will change if we change its environment—how it will alter its prices if a sales tax is levied on its products. We can sometimes make this prediction—and economists do make it repeatedly—without detailed assumptions about the adaptive mechanism, the decision-making apparatus that constitutes the inner environment of the business firm.

Thus the first advantage of dividing outer from inner environment in studying an adaptive or artificial system is that we can often predict behavior from knowledge of the system's goals and its outer environment, with only minimal assumptions about the inner environment. An instant corollary is that we often find quite different inner environments accomplishing identical or similar goals in identical or similar outer environments—airplanes and birds, dolphins and tunafish, weight-driven clocks and battery-driven clocks, electrical relays and transistors.

There is often a corresponding advantage in the division from the standpoint of the inner environment. In very many cases whether a particular system will achieve a particular goal or adaptation depends on only a few characteristics of the outer environment and not at all on the detail of that environment. Biologists are familiar with this property of adaptive systems under the label of homeostasis. It is an important property of most good designs, whether biological or artifactual. In one way or another the designer insulates the inner system from the environment, so that an invariant relation is maintained between inner system and goal, independent of variations over a wide range in most parameters that characterize the outer environment. The ship's chronometer reacts to the pitching of the ship only in the negative sense of maintaining an invariant relation of the hands on its dial to the real time, independently of the ship's motions.

Quasi independence from the outer environment may be maintained by various forms of passive insulation, by reactive negative feedback (the most frequently discussed form of insulation), by predictive adaptation, or by various combinations of these.

Functional Description and Synthesis

In the best of all possible worlds—at least for a designer—we might even hope to combine the two sets of advantages we have described that derive from factoring an adaptive system into goals, outer environment, and inner environment. We might hope to be able to characterize the main properties of the system and its behavior without elaborating the detail of *either* the outer or inner environments. We might look toward a science of the artificial that would depend on the relative simplicity of the interface as its primary source of abstraction and generality.

Consider the design of a physical device to serve as a counter. If we want the device to be able to count up to one thousand, say, it must be capable of assuming any one of at least a thousand states, of maintaining itself in any given state, and of shifting from any state to the "next" state. There are dozens of different inner environments that might be used (and have been used) for such a device. A wheel notched at each twenty minutes of arc, and with a ratchet device to turn and hold it, would do the trick. So would a string of ten electrical switches properly connected to represent binary numbers. Today instead of switches we are likely to use transistors or other solid-state devices.[5]

Our counter would be activated by some kind of pulse, mechanical or electrical, as appropriate, from the outer environment. But by building an appropriate transducer between the two environments, the physical character of the interior pulse could again be made independent of the physical character of the exterior pulse—the counter could be made to count anything.

Description of an artifice in terms of its organization and functioning—its interface between inner and outer environments—is a major objective of invention and design activity. Engineers will find familiar the language of the following claim quoted from a 1919 patent on an improved motor controller:

> What I claim as new and desire to secure by Letters Patent is:
> 1 In a motor controller, in combination, reversing means, normally effective field-weakening means and means associated with said reversing means for

---

[5] The theory of functional equivalence of computing machines has had considerable development in recent years. See Marvin L. Minsky, *Computation: Finite and Infinite Machines* (Englewood Cliffs, N.J.: Prentice-Hall, 1967), chapters 1–4.

> rendering said field-weakening means ineffective during motor starting and thereafter effective to different degrees determinable by the setting of said reversing means . . .[6]

Apart from the fact that we know the invention relates to control of an electric motor, there is almost no reference here to specific, concrete objects or phenomena. There is reference rather to "reversing means" and "field-weakening means," whose further purpose is made clear in a paragraph preceding the patent claims:

> The advantages of the special type of motor illustrated and the control thereof will be readily understood by those skilled in the art. Among such advantages may be mentioned the provision of a high starting torque and the provision for quick reversals of the motor.[7]

Now let us suppose that the motor in question is incorporated in a planing machine (see figure 2). The inventor describes its behavior thus:

> Referring now to [figure 2], the controller is illustrated in outline connection with a planer (100) operated by a motor M, the controller being adapted to govern the motor M and to be automatically operated by the reciprocating bed (101) of the planer. The master shaft of the controller is provided with a lever (102) connected by a link (103) to a lever (104) mounted upon the planer frame and projecting into the path of lugs (105) and (106) on the planer bed. As will be understood, the arrangement is such that reverse movements of the planer bed will, through the connections described, throw the master shaft of the controller back and forth between its extreme positions and in consequence effect selective operation of the reversing switches (1) and (2) and automatic operation of the other switches in the manner above set forth.[8]

In this manner the properties with which the inner environment has been endowed are placed at the service of the goals in the context of the outer environment. The motor will reverse periodically under the control of the position of the planer bed. The "shape" of its behavior—the time path, say, of a variable associated with the motor—will be a function of the "shape" of the external environment—the distance, in this case, between the lugs on the planer bed.

The device we have just described illustrates in microcosm the nature of artifacts. Central to their description are the goals that link the inner

---

[6] U.S. Patent 1,307,836, granted to Arthur Simon, June 24, 1919.

[7] Ibid.

[8] Ibid.

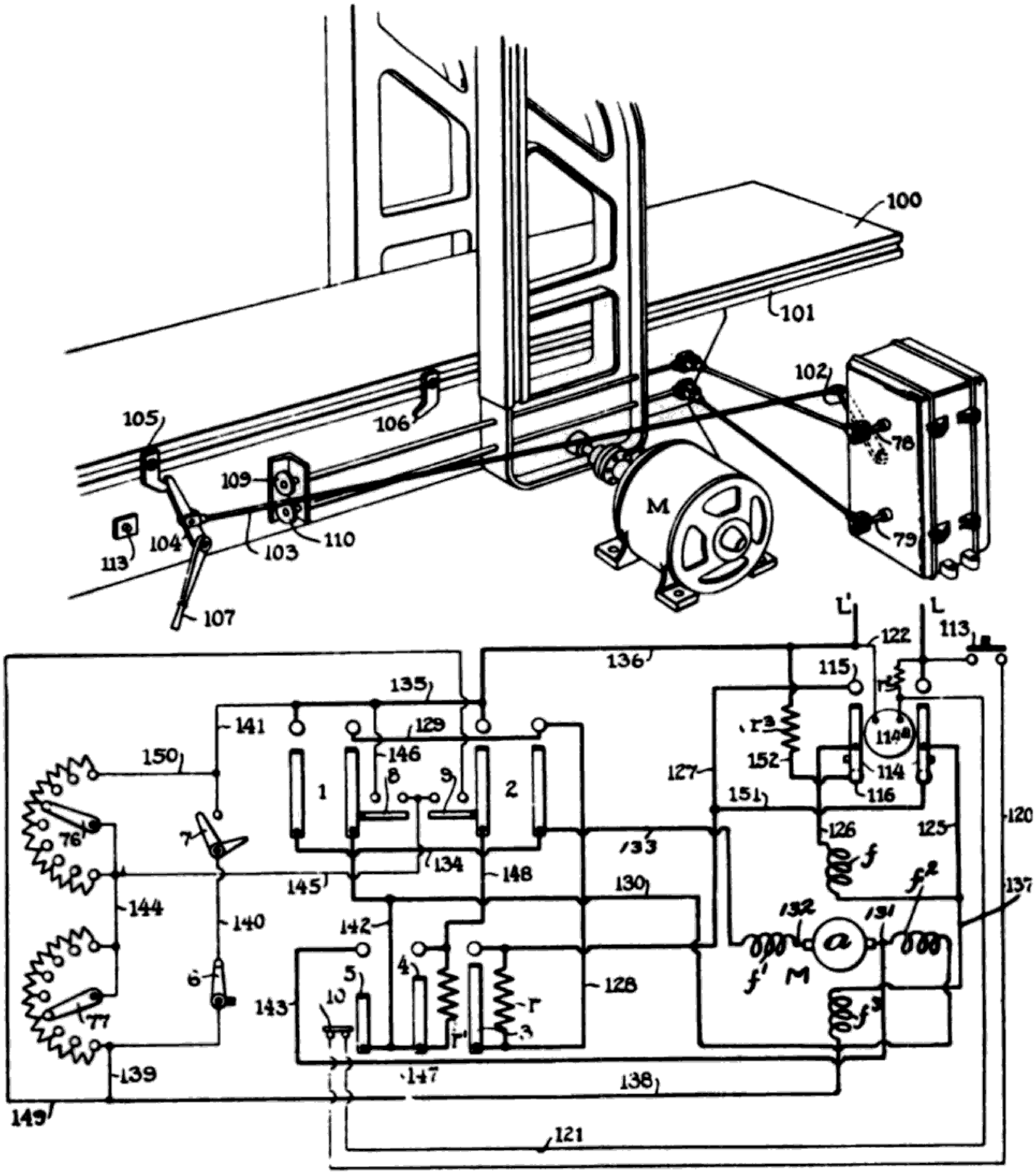

Figure 2
Illustrations from a patent for a motor controller

to the outer system. The inner system is an organization of natural phenomena capable of attaining the goals in some range of environments, but ordinarily there will be many functionally equivalent natural systems capable of doing this.

The outer environment determines the conditions for goal attainment. If the inner system is properly designed, it will be adapted to the outer environment, so that its behavior will be determined in large part by the behavior of the latter, exactly as in the case of "economic man." To predict how it will behave, we need only ask, "How would a rationally designed system behave under these circumstances?" The behavior takes on the shape of the task environment.[9]

Limits of Adaptation

But matters must be just a little more complicated than this account suggests. "If wishes were horses, all beggars would ride." And if we could always specify a protean inner system that would take on exactly the shape of the task environment, designing would be synonymous with wishing. "Means for scratching diamonds" defines a design objective, an objective that *might* be attained with the use of many different substances. But the design has not been achieved until we have discovered at least one realizable inner system obeying the ordinary natural laws—one material, in this case, hard enough to scratch diamonds.

Often we shall have to be satisfied with meeting the design objectives only approximately. Then the properties of the inner system will "show through." That is, the behavior of the system will only partly respond to the task environment; partly, it will respond to the limiting properties of the inner system.

Thus the motor controls described earlier are aimed at providing for "quick" reversal of the motor. But the motor must obey electromagnetic and mechanical laws, and we could easily confront the system with a task where the environment called for quicker reversal than the motor was capable of. In a benign environment we would learn from the motor only what it had been called upon to do; in a taxing environment we would learn something about its internal structure—specifically about those aspects of the internal structure that were chiefly instrumental in limiting performance.[10]

---

[9] On the crucial role of adaptation or rationality—and their limits—for economics and organization theory, see the introduction to part IV, "Rationality and Administrative Decision Making," of my *Models of Man* (New York: Wiley, 1957); pp. 38–41, 80–81, and 240–244 of *Administrative Behavior;* and chapter 2 of this book.

[10] Compare the corresponding proposition on the design of administrative organizations: "Rationality, then, does not determine behavior. Within the area of rationality behavior is perfectly flexible and adaptable to abilities, goals, and

There are two related ways in which simulation can provide new knowledge—one of them obvious, the other perhaps a bit subtle. The obvious point is that, even when we have correct premises, it may be very difficult to discover what they imply. All correct reasoning is a grand system of tautologies, but only God can make direct use of that fact. The rest of us must painstakingly and fallibly tease out the consequences of our assumptions.

Thus we might expect simulation to be a powerful technique for deriving, from our knowledge of the mechanisms governing the behavior of gases, a theory of the weather and a means of weather prediction. Indeed, as many people are aware, attempts have been under way for some years to apply this technique. Greatly oversimplified, the idea is that we already know the correct basic assumptions, the local atmospheric equations, but we need the computer to work out the implications of the interactions of vast numbers of variables starting from complicated initial conditions. This is simply an extrapolation to the scale of modern computers of the idea we use when we solve two simultaneous equations by algebra.

This approach to simulation has numerous applications to engineering design. For it is typical of many kinds of design problems that the inner system consists of components whose fundamental laws of behavior—mechanical, electrical, or chemical—are well known. The difficulty of the design problem often resides in predicting how an assemblage of such components will behave.

Simulation of Poorly Understood Systems

The more interesting and subtle question is whether simulation can be of any help to us when we do not know very much initially about the natural laws that govern the behavior of the inner system. Let me show why this question must also be answered in the affirmative.

First, I shall make a preliminary comment that simplifies matters: we are seldom interested in explaining or predicting phenomena in all their particularity; we are usually interested only in a few properties abstracted from the complex reality. Thus, a NASA-launched satellite is surely an artificial object, but we usually do not think of it as "simulating" the moon or a planet. It simply obeys the same laws of physics, which relate only to its inertial and gravitational mass, abstracted from most of its other properties. It *is* a moon. Similarly electric energy that entered my house from the early atomic generating station at Shippingport did not "simulate" energy generated by means of a coal plant or a windmill. Maxwell's equations hold for both.

The more we are willing to abstract from the detail of a set of phenomena, the easier it becomes to simulate the phenomena. Moreover we do not have to know, or guess at, all the internal structure of the system but only that part of it that is crucial to the abstraction.

It is fortunate that this is so, for if it were not, the topdown strategy that built the natural sciences over the past three centuries would have been infeasible. We knew a great deal about the gross physical and chemical behavior of matter before we had a knowledge of molecules, a great deal about molecular chemistry before we had an atomic theory, and a great deal about atoms before we had any theory of elementary particles—if indeed we have such a theory today.

This skyhook-skyscraper construction of science from the roof down to the yet unconstructed foundations was possible because the behavior of the system at each level depended on only a very approximate, simplified, abstracted characterization of the system at the level next beneath.[13] This is lucky, else the safety of bridges and airplanes might depend on the correctness of the "Eightfold Way" of looking at elementary particles.

Artificial systems and adaptive systems have properties that make them particularly susceptible to simulation via simplified models. The characterization of such systems in the previous section of this chapter

---

[13] This point is developed more fully in "The Architecture of Complexity," chapter 8 in this volume. More than fifty years ago, Bertrand Russell made the same point about the architecture of mathematics. See the "Preface" to *Principia Mathematica:* ". . . the chief reason in favour of any theory on the principles of mathematics must always be inductive, i.e., it must lie in the fact that the theory in question enables us to deduce ordinary mathematics. In mathematics, the greatest degree of self-evidence is usually not to be found quite at the beginning, but at some later point; hence the early deductions, until they reach this point, give reasons rather for believing the premises because true consequences follow from them, than for believing the consequences because they follow from the premises." Contemporary preferences for deductive formalisms frequently blind us to this important fact, which is no less true today than it was in 1910.

The Sciences of the Artificial

Third Edition

Herbert A. Simon

Herbert A. Simon pioneered the study of the artificial. The natural sciences describe "natural" objects and phenomena. The sciences of the artificial describe objects and phenomena—artifacts—that result from human intervention in the natural world. Much of our daily world is artificial—from the climate-controlled air we breathe to the automobiles we drive and the laws that tell us how fast we may drive them. Aimed at satisfying human purposes, artifacts are not exempt from natural law but are adapted to the environments in which they operate. Conceived in the human activity called design, many are of immense complexity; computers, for example, are invaluable in studying natural as well as artificial complexity.

The first edition of *The Sciences of the Artificial*, published in 1969, quickly became a classic for its insights into complex systems in general, the process of design, and artificial intelligence and its contribution to our understanding of human intelligence. Simon's fundamental theses are as relevant today as when he began his investigations. His examination of complex systems has found important applications in the cognitive, political, economic, and biological sciences, among others. In particular, Simon has proposed new methods for the study of design—devising artifacts to attain goals.

Today complexity remains a topic of intense interest. A new chapter in this third edition of *The Sciences of the Artificial* discusses current themes and the tools—the mathematics of chaos, adaptive systems, genetic algorithms—used to analyze complexity and complex systems. Throughout the new edition, the author integrates recent advances in cognitive psychology and the science of design, confirming and extending the book's basic thesis: that a physical symbol system has the necessary means for intelligent action. The revised chapter on "Economic Reality" reflects new thinking and research on the roles of organizations and markets in economic systems.

Herbert A. Simon is Richard King Mellon University Professor of Computer Science and Psychology at Carnegie Mellon University and the 1978 Nobel Laureate in Economics.

The MIT Press | Massachusetts Institute of Technology | Cambridge, Massachusetts | 02142
www-mitpress.mit.edu

Cover image © 1996 PhotoDisc, Inc.

90000
9 780262 691918

0-262-69191-4