The Sciences of the Artificial

Simon, Herbert A.

Excerpt from ch1. Understanding the Natural and the Artificial Worlds (1996), The Sciences of the Artificial (3rd ed.). Cambridge, MA: MIT Press. ISBN 9780262193740.

DOI: 10.36253/contest-17038

This excerpt from the first chapter of The Sciences of the Artificial (1969; 1996) by Herbert A. Simon establishes the epistemological foundations for distinguishing natural sciences from the "sciences of the artificial." While natural sciences seek hidden patterns to explain how things are, the sciences of the artificial deal with objects synthesised by human beings, characterised by functions, goals, and normative imperatives—addressing how things "ought" to be. Simon introduces the crucial concept of the artefact as an "interface" between an "inner" environment (the substance and organisation of the object itself) and an "outer" environment (the context in which it operates); the artefact's effectiveness depends on the successful adaptation of these two environments to one another. The text further explores

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

the role of simulation as a source of new knowledge that can reveal the hidden implications of known premises. It defines both computers and the human mind as "physical symbol systems." According to Simon, intelligence is fundamentally the work of these systems, which can encode information, manipulate structures, and adapt to their environment.

The re-proposal of this classic text within the contemporary context of urban studies and Artificial Intelligence (PlanAIr) is driven by three fundamental reasons. First, Simon provides a critical ontological definition, reminding us that the world we inhabit is predominantly man-made. In this view, the city is the artefact par excellence: not a natural phenomenon to be passively observed, but a complex, designed system that must answer to human purposes, thus legitimising urban planning as a rigorous science of the artificial. Second, the vision of the artefact as a "meeting point" between inner and outer environments offers a powerful metaphor for Urban AI. Intelligent technologies in the city act as an interface between physical infrastructure and citizens' social or environmental dynamics, requiring mutual adaptation to function effectively. Finally, as Simon's work is foundational to symbolic Artificial Intelligence, revisiting it today allows us to grasp the theoretical roots of rule-based and logical AI. This historical perspective is crucial for distinguishing and potentially integrating symbolic approaches with the currently dominant data-driven paradigms, thereby recovering the capacity to reason about goals, meanings, and design imperatives rather than relying solely on raw data processing.

In this article, we examine these ontological issues, discuss existing frameworks that aim to unify fragmented information, and explore the practical implications for urban AI applications. The thesis is that ontologies—structured and formal representations of knowledge—offer a powerful tool to address the challenges outlined above, while serving as a blueprint for defining, categorizing, and interrelating the entities present in urban environments and putting them to work in urban planning.

> 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 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 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 [...]

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[2]?

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 Artifact As "Interface"

We can view the matter quite symmetrically. An artifact can be thought of as a meeting point an "interface" in today's terms between an "inner" environment, the substance and organization of the artifact itself, and an ''outer" environment, the surroundings in which it operates. If the inner environment is appropriate to the outer environment, or vice versa, the artifact will serve its intended purpose. Thus, if the clock is immune to buffeting, it will serve as a ship's chronometer. (And conversely, if it isn't, we may salvage it by mounting it on the mantel at home.)

Notice that this way of viewing artifacts applies equally well to many things that are not man-made to all things in fact that can be regarded as adapted to some situation; and in particular it applies to the living systems

that have evolved through the forces of organic evolution. A theory of the airplane draws on natural science for an explanation of its inner environment (the power plant, for example), its outer environment (the character of the atmosphere at different altitudes), and the relation between its inner and outer environments (the movement of an air foil through a gas). But a theory of the bird can be divided up in exactly the same way.

Given an airplane, or given a bird, we can analyze them by the methods of natural science without any particular attention to purpose or adaptation, without reference to the interface between what I have called the inner and outer environments. After all, their behavior is governed by natural law just as fully as the behavior of anything else (or at least we all believe this about the airplane, and most of us believe it about the bird).

Understanding by Simulating

Artificiality connotes perceptual similarity but essential difference, resemblance from without rather than within. In the terms of the previous section we may say that the artificial object imitates the real by turning the same face to the outer system, by adapting, relative to the same goals, to comparable ranges of external tasks. Imitation is possible because distinct physical systems can be organized to exhibit nearly identical behavior. The damped spring and the damped circuit obey the same second-order linear differential equation; hence we may use either one to imitate the other.

Techniques of Simulation

Because of its abstract character and its symbol manipulating generality, the digital computer has greatly extended the range of systems whose behavior can be imitated. Generally we now call the imitation "simulation," and we try to understand the imitated system by testing the simulation in a variety of simulated, or imitated, environments.

Simulation, as a technique for achieving understanding and predicting the behavior of systems, predates of course the digital computer. The model basin and the wind tunnel are valued means for studying the behavior of large systems by modeling them in the small, and it is quite certain that Ohm's law was suggested to its discoverer by its analogy with simple hydraulic phenomena.

Simulation may even take the form of a thought experiment, never actually implemented dynamically. One of my vivid memories of the Great Depression is of a large multi colored chart in my father's study that represented a hydraulic model of an economic system (with different fluids for money and goods). The chart was devised by a technocratically inclined engineer named Dahlberg. The model never got beyond the pen-and-paint stage at that time, but it could be used to trace through the imputed consequences of particular economic

measures or events provided the theory was right![3] [...]

As my formal education in economics progressed, I acquired a disdain for that naive simulation, only to discover after World War II that a distinguished economist, Professor A. W. Phillips had actually built the Moniac, a hydraulic model that simulated a Keynesian economy [...].[4] Of course Professor Phillips's simulation incorporated a more nearly correct theory than the earlier one and was actually constructed and operated two points in its favor. However, the Moniac, while useful as a teaching tool, told us nothing that could not be extracted readily from simple mathematical versions of Keynesian theory and was soon priced out of the market by the growing number of computer simulations of the economy.

Simulation As a Source of New Knowledge

This brings me to the crucial question about simulation: How can a simulation ever tell us anything that we do not already know? The usual implication of the question is that it can't. As a matter of fact, there is an interesting parallelism, which I shall exploit presently, between two assertions about computers and simulation that one hears frequently:

1. A simulation is no better than the assumptions built into it.
2. A computer can do only what it is programmed to do.

I shall not deny either assertion, for both seem to me to be true. But despite both assertions simulation can tell us things we do not already know.

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.
[...]

The Computer As Artifact

No artifact devised by man is so convenient for this kind of functional description as a digital computer. It is truly protean, for almost the only ones of its properties that are detectable in its behavior (when it is operating properly!) are the organizational properties. The speed with which it performs it basic operations may allow us to infer a little about its physical components and their natural laws; speed data, for example, would allow us to rule out certain kinds of "slow" components. For the rest, almost no interesting statement that one can make about an operating computer bears any particular relation to the specific nature of the hardware. A computer is an organization of elementary functional components in which, to a high approximation, only the function Page 18 performed by those components is relevant to the behavior of the whole system.[5]

Computers As Abstract Objects

This highly abstractive quality of computers makes it easy to introduce mathematics into the study of their theory and has led some to the erroneous conclusion that, as a computer science emerges, it will necessarily be a mathematical rather than an empirical science. Let me take up these two points in turn: the relevance of mathematics to computers and the possibility of studying computers empirically.

Some important theorizing, initiated by John von Neumann, has been done on the topic of computer reliability. The question is how to build a reliable system from unreliable parts. Notice that this is not posed as a question of physics or physical engineering. The components engineer is assumed to have done his best, but the parts are still unreliable! We can cope with the unreliability only by our manner of organizing them.

To turn this into a meaningful problem, we have to say a little more about the nature of the unreliable parts. Here we are aided by the knowledge that any computer can be assembled out of a small array of simple, basic elements. For instance, we may take as our primitives the so-called Pitts-McCulloch neurons. As their name implies, these components were devised in analogy to the supposed anatomical and functional characteristics of neurons in the brain, but they are highly abstracted. They are formally isomorphic with the simplest kinds of switching circuits "and" "or," and "not'' circuits. We postulate, now, that we are to build a system from such elements and that each elementary part has a specified probability of

functioning correctly. The problem is to arrange the elements and their interconnections in such a way that the complete system will perform reliably.

The important point for our present discussion is that the parts could as well be neurons as relays, as well relays as transistors. The natural laws governing relays are very well known, while the natural laws governinneurons are known most imperfectly. But that does not matter, for all that is relevant for the theory is that the components have the specified level of unreliability and be interconnected in the specified way.

This example shows that the possibility of building a mathematical theory of a system or of simulating that system does not depend on having an adequate micro theory of the natural laws that govern the system components. Such a micro theory might indeed be simply irrelevant.

Computers As Empirical Objects

We turn next to the feasibility of an empirical science of computers as distinct from the solid-state physics or physiology of their componentry.[6] As a matter of empirical fact almost all of the computers that have been designed have certain common organizational features. They almost all can be decomposed into an active processor (Babbage's "Mill") and a memory (Babbage's "Store") in combination with input and output devices. (Some of the larger systems, somewhat in the manner of colonial algae, are assemblages of smaller systems having some or all of these components. But perhaps I may oversimplify for the moment.) They are all capable of storing symbols (program) that can be interpreted by a program-control component and executed. Almost all have exceedingly limited capacity for simultaneous, parallel activity they are basically one-thing-at-a-time systems. Symbols generally have to be moved from the larger memory components into the central processor before they can be acted upon. The systems are capable of only simple basic actions: recoding symbols, storing symbols, copying symbols, moving symbols, erasing symbols, and comparing symbols.

Since there are now many such devices in the world, and since the properties that describe them also appear to be shared by the human central nervous system, nothing prevents us from developing a natural history of them. We can study them as we would rabbits or chipmunks and discover how they behave under different patterns of environmental stimulation. Insofar as their behavior reflects largely the broad functional. The research that was done to design computer time-sharing systems is a good example of the study of computer behavior as an empirical phenomenon. Only fragments of theory were available to guide the design of a time-sharing system or to predict how a system of

a specified design would actually behave in an environment of users who placed their several demands upon it. Most actual designs turned out initially to exhibit serious deficiencies, and most predictions of performance were startlingly inaccurate.

Under these circumstances the main route open to the development and improvement of time-sharing systems was to build them and see how they behaved. And this is what was done. They were built, modified, and improved in successive stages. Perhaps theory could have anticipated these experiments and made them unnecessary. In fact it didn't, and I don't know anyone intimately acquainted with these exceedingly complex systems who has very specific ideas as to how it might have done so. To understand them, the systems had to be constructed, and their behavior observed.[7]

In a similar vein computer programs designed to play games or to discover proofs for mathematical theorems spend their lives in exceedingly large and complex task environments. Even when the programs themselves are only moderately large and intricate (compared, say, with the monitor and operating systems of large computers), too little is known about their task environments to permit accurate prediction of how well they will perform, how selectively they will be able to search for problem solutions.

Here again theoretical analysis must be accompanied by large amounts of experimental work. A growing literature reporting these experiments is beginning to give us precise knowledge about the degree of heuristic power of particular heuristic devices in reducing the size of the problem spaces that must be searched. In theorem proving, for example, there has been a whole series of advances in heuristic power based on and guided by empirical exploration: the use of the Herbrand theorem, the resolution principle, the set-of-support principle, and so on [...]. [8]

Computers and Thought

As we succeed in broadening and deepening our knowledge theoretical and empirical about computers, we discover that in large part their behavior is governed by simple general laws, that what appeared as complexity in the computer program was to a considerable extent complexity of the environment to which the program was seeking to adapt its behavior. This relation of program to environment opened up an exceedingly important role for computer simulation as a tool for achieving a deeper understanding of human behavior. For if it is the organization of components, and not their physical properties, that largely determines behavior, and if computers are organized somewhat in the image of man, then the computer becomes an obvious device for exploring the consequences of alternative organizational assumptions for human behavior. Psychology could move forward

without awaiting the solutions by neurology of the problems of component design however interesting and significant these components turn out to be.

Symbol Systems: Rational Artifacts

The computer is a member of an important family of artifacts called symbol systems, or more explicitly, physical symbol systems [...].[9] Another important member of the family (some of us think, anthropomorphically, it is the most important) is the human mind and brain. It is with this family of artifacts, and particularly the human version of it, that we will be primarily concerned in this book. Symbol systems are almost the quintessential artifacts, for adaptivity to an environment is their whole raison d'être. They are goalseeking, information-processing systems, usually enlisted in the service of the larger systems in which they are incorporated.

Basic Capabilities of Symbol Systems

A physical symbol system holds a set of entities, called symbols. These are physical patterns (e.g., chalk marks on a blackboard) that can occur as components of symbol structures (sometimes called "expressions"). As I have already pointed out in the case of computers, a symbol system also possesses a number of simple processes that operate upon symbol structures processes that create, modify, copy, and destroy symbols. A physical symbol system is a machine that, as it moves through time, produces an evolving collection of symbol structures.[10] Symbol structures can, and commonly do, serve as internal representations (e.g., "mental images") of the environments to which the symbol system is seeking to adapt. They allow it to model that environment with greater or less veridicality and in greater or less detail, and consequently to reason about it. Of course, for this capability to be of any use to the symbol system, it must have windows on the world and hands, too. It must have means for acquiring information from the external environment that can be encoded into internal symbols, as well as means for producing symbols that initiate action upon the environment. Thus it must use symbols to designate objects and relations and actions in the world external to the system.

Symbols may also designate processes that the symbol system can interpret and execute. Hence the programs that govern the behavior of a symbol system can be stored, along with other symbol structures, in the system's own memory, and executed when activated.

Symbol systems are called "physical" to remind the reader that they exist as real-world devices, fabricated of glass and metal (computers) or flesh and blood (brains). In the past we have been more accustomed to thinking of the symbol systems of mathematics and logic as abstract and disembodied, leaving out of

account the paper and pencil and human minds that were required actually to bring them to life. Computers have transported symbol systems from the platonic heaven of ideas to the empirical world of actual processes carried out by machines or brains, or by the two of them working together.

Intelligence As Computation

The three chapters that follow rest squarely on the hypothesis that intelligence is the work of symbol systems. Stated a little more formally, the hypothesis is that a physical symbol system of the sort I have just described has the necessary and sufficient means for general intelligent action.

The hypothesis is clearly an empirical one, to be judged true or false on the basis of evidence. One task of chapters 3 and 4 will be to review some of the evidence, which is of two basic kinds. On the one hand, by constructing computer programs that are demonstrably capable of intelligent action, we provide evidence on the sufficiency side of the hypothesis. On the other hand, by collecting experimental data on human thinking that tend to show that the human brain operates as a symbol system, we add plausibility to the claims for necessity, for such data imply that all known intelligent systems (brains and computers) are symbol systems [...].

Notes

[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.
[2] 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.
[3] For some published versions of this model, see A. O. Dahlberg, National Income Visualized (N.Y.: Columbia University Press, 1956).
[4] A. W. Phillips, "Mechanical Models in Economic Dynamics," Economica, New Series, 17 (1950):283 305.
[5] On the subject of this and the following paragraphs, see M. L. Minsky, op. cit.; then John von Neumann, "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components," in C. E. Shannon and J. McCarthy (eds.), Automata Studies (Princeton: Princeton University Press, 1956).
[6] A. Newell and H. A. Simon, "Computer Science as Empirical Inquiry," Communications of the ACM, 19(March 1976):113 126. See also H. A. Simon, "Artificial Intelligence: An Empirical Science," Artificial Intelligence, 77(1995):95 127.
[7] The empirical, exploratory flavor of computer research is nicely captured by the account of Maurice V. Wilkes in his 1967 Turing Lecture, "Computers Then and Now," Journal of the Association for Computing Machinery, 15(January 1968):1 7.
[8] Note, for example, the empirical data in Lawrence Wos, George A. Robinson, Daniel F. Carson, and Leon Shalla, "The Concept of Demodulation in Theorem Proving," Journal of the Association for Computing Machinery, 14(October 1967):698 709, and in several of the earlier papers referenced there. See also the collection of programs in Edward Feigenbaum and Julian Feldman (eds.), Computers and Thought (New York: McGraw-Hill, 1963). It is common practice in the field to title papers about heuristic programs, "Experiments with an XYZ Program".
[9] In the literature the phrase information-processing system is used more frequently than symbol system. I will use the two terms as synonyms.
[10] Newell and Simon, "Computer Science as Empirical Inquiry," p. 116.