Scientific Methods in Computer Science

Article · December 2002

Source: CiteSeer

1 author:

Gordana Dodig Crnkovic
Chalmers University of Technology

Scientific Methods in Computer Science

Gordana Dodig-Crnkovic
Department of Computer Science
Mälardalen University
Västerås, Sweden
+46 21 15 17 25
gordana.dodig-crnkovic@mdh.se

ABSTRACT
This paper analyzes scientific aspects of Computer Science. First it defines science and scientific method in general. It gives a discussion of relations between science, research, development and technology.

The existing theory of science (Popper, Carnap, Kuhn, Chalmers) has *Physics* as an ideal. Not many sciences come close to that ideal. Philosophy of Science (Theory of Science) as it is today is not of much help when trying to analyze Computer Science.

Computer Science is a new field and its object of investigation (universe) is a computer, which is an ever-developing artifact, the materialization of the ideas that try to structure knowledge and the information about the world, including computers themselves.

However different, Computer Science has its basis in Logic and Mathematics, and both theoretical and experimental research methods follow patterns of classical scientific fields. Computer modeling and simulation as a method is specific for the discipline, and it is going to develop even more in the future, not only applied to computers, but also to other scientific as well as commercial and artistic fields.

Keywords
Computer Science, Theory of science, scientific methodology.

1. INTRODUCTION
It is not so obvious, as the name might suggest, that the Computer Science qualifies as "science". Computer Science (CS) is a young discipline and necessarily starting from the outset very different from Mathematics, Physic and similar "classic" sciences, that all have their origins in the philosophy of ancient Greece.

Emerging in modern time (in 1940's the first electronic digital computer was built), CS has necessarily other already existing sciences in the background.

Computer Science draws its foundations from a wide variety of disciplines. Study of Computer Science consequently requires utilizing concepts from many different fields. Computer Science integrates theory and practice, abstraction (general) and design (specific).

The historical development has led to emergence of a big number of sciences that communicate more and more not only because the means of communication are getting very convenient and effective, but also because a need increases for getting a holistic view of our world, that is presently strongly dominated by reductionism.

2. WHAT IS SCIENCE

The whole is more than the sum of its parts.

Aristotle, Metaphysica

2.1 Classical Sciences
Talking about "Science" we actually mean plurality of different sciences. Different sciences differ very much from each other.

The definition of science is therefore neither simple nor unambiguous. See [1] and [2] for several possible classifications. For example, history and linguistics are often but not always catalogued as sciences.

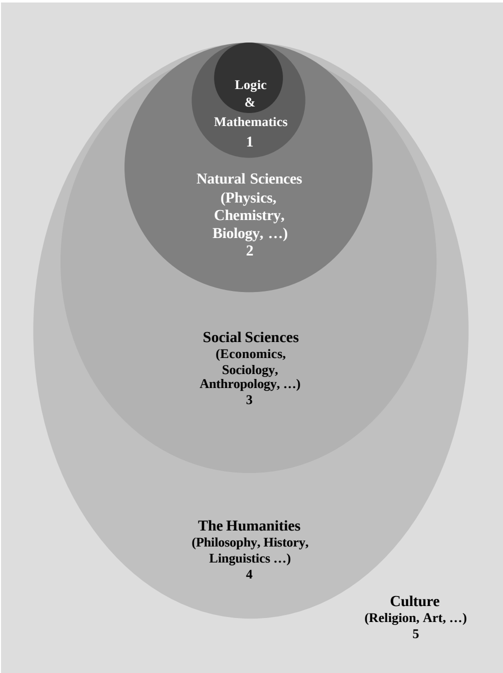

Figure 1 What is science? One possible view.

From the chosen scheme of the figure above we can realize that sciences have *specific areas of validity*. The Logic and Mathematics (the most abstract and at the same time the most exact sciences) are more or less important part of every other science. They are very essential for Physics, less important for Chemistry and Biology, and their significance continues to decrease towards the outer regions of our scheme.

The logical reasoning as a basis of all human knowledge is of course present in every kind of science as well as in philosophy.

The structure of Figure 1 may be seen in analogy with looking into a microscope. With the highest resolution we can reach the innermost region. Inside the central region Logic is not only the tool used to make conclusions. It is at the same time the *object of investigation*. Even though big parts of Mathematics can be reduced to Logic (Frege, Rusell and Whitehead) the complete reduction is impossible.

On every step of zooming out, the inner regions are given as prerequisites for the outer ones. Physics is using Mathematics and Logic as tools, without questioning their internal structure. In that way information about the deeper structure of Mathematics and Logic is hidden looking from the outside. In much the same way, Physics is a prerequisite for Chemistry that is a hidden level inside Biology etc.

The basic idea of Figure 1 is to show in a schematic way the relation between the three main groups of sciences (Logic & Mathematics, Natural Sciences and Social Sciences) as well as the connections to thought systems represented by the Humanities.

Finally the whole body of human knowledge, scientific and speculative is immersed in and impregnated by the cultural environment.

Table 1 Sciences, objects and methods

SCIENCE	OBJECTS	DOMINATING METHOD
	Simple	Reductionism (analysis)
Logic & Mathematics	Abstract objects: propositions, numbers, ...	Deduction
Natural Sciences	Natural objects: physical bodies, fields and interactions, living organisms ...	Hypothetico-deductive method
Social Sciences	Social objects: human individuals, groups, society,	Hypothetico-deductive method + Hermeneutics
Humanities	Cultural objects: human ideas, actions and relationships, language, artifacts…	Hermeneutics
	Complex	Holism (synthesis)

The innermost sciences, *Logic and Mathematics* are the most fundamental ones and the ones with the highest degree of certainty. They have the most abstract and the simplest objects of investigation. Their language is the most formal one. They rely predominantly on the deductive method. It is however important to notice that *the basic elements in both Logic and Mathematics have been extracted from our real-life language and purified into set of well defined formulae/symbolic expressions via an essentially inductive process*.

The next region, *Natural Sciences*, is not an axiomatized theory as the previous one. Physics, which is the ideal of science for many philosophers of science (Popper, Carnap, Kuhn, Chalmers [4-7]) contains both theoretical parts with pure mathematical formulations derived from "first principles" and parts that are empirical i.e. short-cut expressions for observed facts that are built-in into system as they proved useful. Far away from all theoretical Physics can be axiomatized. Even less so is the case for experimental Physics, for quite obvious reasons.

Natural sciences are dominated by a method that Popper calls *hypotetico-deductive method*.

Social Sciences include sociology, pedagogic, anthropology, economics etc. The objects studied are humans as social beings, alone or in a group. Social sciences primarily rely on the qualitative methods. The aim is to *understand (in the sense of hermeneutics)* and *describe* phenomena. The quantitative aspects of their methodology are related to statistical methods.

The Humanities (The Liberal Arts) include philosophy, history, linguistics and similar. The difference between Humanities and Social Science is not a very sharp one, but we can say that Humanities predominantly have a qualitative approach, and very rarely depend on any statistical methods.

Figure 1 represents a dynamic scheme seen in a specific moment. For example a corresponding scheme for the medieval sciences would be very different. In other words, *culture* is like a flow that all sciences follow. Albeit very slow, that flow steadily changes the framework for all the sciences.

2.2 Sciences Belonging to Several Fields

The development of human thought parallel to the development of human society has led to an emergence of sciences that do not belong to any of the classic types we have described earlier (see Figure 1), but rather share common parts with several of these.

Many of the modern sciences are of interdisciplinary, "eclectic" type. It is a trend for new sciences to search their methods and even questions in very broad areas. It can be seen as a result of the fact that the communications across the borders of different scientific fields is nowadays much easier and more intense than before.

Computer Science for example includes the field of artificial intelligence that has its roots in mathematical Logic and Mathematics but uses Physics, Chemistry and Biology and even has parts where medicine and psychology are very important.

We seem to be witnessing an exciting paradigm shift:

We should, by the way, be prepared for some radical, and perhaps surprising, transformations of the disciplinary structure of science (technology included) as information processing pervades it. In particular, as we become more aware of the detailed information processes that go on in doing science, the sciences will find themselves increasingly taking a meta-position, in which doing science (observing, experimenting, theorizing, testing, archiving,) will involve understanding these information processes, and building systems that do the object-level science. Then the boundaries between the enterprise of science as a whole (the acquisition and organization of knowledge of the world) and AI (the understanding of how knowledge is acquired and organized) will become increasingly fuzzy.

Allen Newell, Artif. Intell. 25 (1985) 3.

Here we can find a potential of the new synthetic (holistic) world-view that is about to emerge in the future.

3. THE SCIENTIFIC METHOD

The scientific method is the logical scheme used by scientists searching for answers to the questions posed within science. Scientific method is used to produce scientific theories, including both scientific meta-theories (theories about theories) as well as the theories used to design the tools for producing theories (instruments, algorithms, etc). The simple version looks something like this (see also Figure 2):

1. Pose the question in the context of existing knowledge (theory & observations). It can be a new question that old theories are capable of answering (usually the case), or the question that calls for formulation of a new theory.
2. Formulate a hypothesis as a tentative answer.
3. Deduce consequences and make predictions.
4. Test the hypothesis in a specific experiment/theory field. The new hypothesis must prove to fit in the existing world-view (1, “normal science”, according to Kuhn).
   In case the hypothesis leads to contradictions and demands a radical change in the existing theoretical background, it has to be tested particularly carefully. *The new hypothesis has to prove fruitful and offer considerable advantages, in order to replace the existing scientific paradigm*. This is called “*scientific revolution”* (Kuhn) and it happens very rarely. As a rule, the loop 2-3-4 is repeated with modifications of the hypothesis until the agreement is obtained, which leads to 5. If major discrepancies are found the process must start from the beginning, 1.
5. When consistency is obtained the hypothesis becomes a theory and provides a coherent set of propositions that define a new class of phenomena or a new theoretical concept. *The results have to be published*. Theory at that stage is subject of process of ”natural selection” among competing theories (6). A theory is then becoming a framework within which observations/theoretical facts are explained and predictions are made. The process can start from the beginning, but the state 1 has changed to include the new theory/improved old theory.

Figure 2 describes very generally *the logical structure of scientific method used in developing new theories*. As the flow diagram suggests, *science is in a state of permanent change and development.*

The one of the most important qualities of science is its *provisional character:* it is subject to continuous re-examination and self-correction.

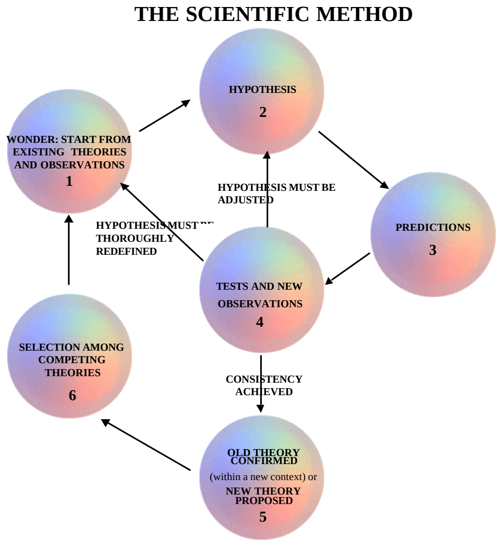

Figure 2 Diagram describing iterative nature of the scientific method (hypothetico-deductive)

It is crucial to understand that the *Logic of science is recursive*. Prior to every observation/experiment/theoretical test it is a hypothesis (2) that has its origins in the pre-existing body of knowledge (1). Every experimental/observational result has a certain world-view built-in. Or, to say it by Feyerabend [8], every experimental data is “theory-contaminated”.

Here it is also interesting to mention that designing new experimental equipment or procedure match the same scheme: (1) Start from existing theoretical/experimental framework; (2) Formulate the problem; (3) Infer consequences; (4) Test if it works as expected; (5-6) Accept.

As soon as a piece of equipment or method is designed and used as a tool for testing new hypotheses, it is supposed that it works according to the design specification. The detailed information about its internal structure is therefore hidden.

The same is true for the existing theoretical context of a theory under development- it is taken for granted.

The scheme of the scientific method in Figure 2 is without a doubt an abstraction and simplification. Critics of the hypothetico-deductive method would argue that there is in fact no such thing as “the scientific method”. *By the term “the scientific method” they actually mean the concrete set of rules defining how to proceed in posing new relevant questions and formulating successful hypotheses.*

The important advantage of the scientific method is that it is *impartial*:[1] one does not have to believe a given researcher, one can (in

[1] *Impartial* is used here as synonymous for *objective, unbiased, unprejudiced, and dispassionate***.** Note, however that this is the statement about *science, not about individual scientists* whose attitude to their pursuit is as a rule passionate**.** The fact that science is shared by the whole scientific community results in theories that are in a great extent *free from individual bias***.** On the other hand the whole of scientific community use to share *com-*

principle) repeat the experiment and determine whether certain results are valid or not. The conclusions will hold irrespective of the state of mind, or the religious persuasion, or the state of consciousness of the investigator. The question of *impartiality* is closely related to *openness* and *universality* of science, which are its fundamental qualities.

A theory is accepted based in the first place on the results obtained through *logical reasoning*, *observations* and/or *experiments*. The results obtained using the scientific method have to be *reproducible*. If the original claims are not verified, the causes of such discrepancies are exhaustively studied.

All scientific truths are *provisional*. But for a hypothesis to get the status of a theory it is necessary to win the confidence of the scientific community. In the fields where there are no commonly accepted theories (as e.g. explanation of the process of creation of the universe- where the "big bang" hypothesis is the most popular one) the number of alternative hypotheses can constitute the body of scientific knowledge.

4. SCIENCE, RESEARCH, TECHNOLOGY

4.1 Aristotle's Science contra Technology

In his famous reflections on science and technology, Aristotle has identified some key distinctions that are still frequently quoted and even used to analyze modern science and technology.

Table 2 Standard distinctions: science vs. technology

	Science	Technology
Object	unchangeable	changeable
Principle of motion	inside	outside
End	knowing the general	knowing the concrete
Activity	theoria: end in itself	poiesis: end in something else
Method	abstraction	modeling concrete (complex)
Process	conceptualizing	optimizing
Innovation form	discovery	invention
Type of result	law-like statements	rule-like statements
Time perspective	long-term	short-term

4.2 Modern Science contra Technology

Traditional sharp binary distinctions between science and technology seem however to fail when applied to contemporary science, because the underlying concepts of science are out-dated. Today's science is much more complex and heterogeneous than science of the Aristotle's time (the contemporary relations are illustrated by Figure 3); the fact that many modern philosophers have difficulty to admit.

That is why philosophy of science is in vital need of a deeper, more realistic understanding of contemporary sciences. The time is ripe for paradigm change in philosophy of science!

---

mon paradigms, which are the very broad concepts deeply rooted in the *culture*. *Paradigm shift is a process that occurs in a very dramatic way, partly because of cultural (not strictly rational) nature of paradigm, (Kuhn).*

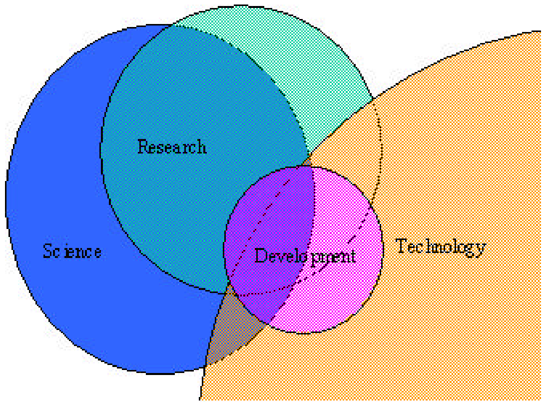

Figure 3 Relations between science, research, development and technology

5. WHAT IS COMPUTER SCIENCE?

Of course it is impossible to give a unique and simple definition of Computer Science. Let me mention some of existing ones:

1. Computer Science is the study of phenomena related to computers, Newell, Perlis and Simon, 1967
2. The discipline of computing is the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application [3], 2001. (Compare to: Computer Science is the study of algorithms, Knuth, 1968)
3. Computer Science is the study of information structures, Wegner, 1968, Curriculum 68
4. Computer Science is the study and management of complexity, Dijkstra, 1969. [9]
5. Computer Science is the mechanization of abstraction, Aho and Ullman1992 [10]

The first definition reflects an *empirical* tradition since it asserts that Computer Science is concerned with the study of a class of phenomena. The second and third definitions reflect a *mathematical* tradition since algorithms and information structures are two abstractions from the phenomena of Computer Science.

The third definition was used by Wegner as the unifying abstraction in his book on Programming Languages, Information Structures and Machine Organization. This view of Computer Science has its historical roots in information theory. It strongly influenced the development of Curriculum 68; a document which has been very prominent in the development of undergraduate Computer Science curricula. It is implicit in the German and French use of the respective terms *"Informatik"* and *"Informatique"* to denote the discipline of Computer Science.

It is interesting to note that the British term "Computer Science" has an empirical orientation, while the corresponding German and French terms have an abstract orientation. This difference in terminology appears to support the view that the nineteenth-century traits of British empiricism and continental abstraction have persisted.

The view that information is the central idea of Computer Science is both scientifically and sociologically suggestive. Scientifically, it suggests a view of Computer Science as a generalization of information theory that is concerned not only with the *transmission of information* but also with its *transformation* and *interpretation*. Sociologically, it suggests an analogy between the industrial revolution, which is concerned with the utilizing of energy, and the computer revolution, which is concerned with the utilizing of information.

The fourth definition reflects the great complexity of engineering problems encountered in managing the construction of complex software-hardware systems.

It is argued in [9] that Computer Science was dominated by empirical research paradigms in the 1950s, by mathematical research paradigms in the 1960s and by engineering oriented paradigms beginning with the 1970s.

The diversity of research paradigms in Computer Science may be responsible for divergences of opinion concerning the nature of Computer Science research.

The fundamental question underlying all computing is: What can be (efficiently) automated?

Computer Science is a field of study that is concerned with theoretical and applied disciplines in the development and use of computers for information storage and processing, Mathematics, Logic, science, and many other areas.

The discipline was born in 1940s through the joining of Mathematical Logic, algorithm theory and electronic computer.

Logic is important not only because it forms the basis of every programming language, or because of its investigating into the limits of automatic calculation, but also because of its insight that strings of symbols (also encoded as numbers) can be interpreted both as data and as programs.

Sub-areas of computing according to [3]:

1. Discrete Structures
2. Programming Fundamentals
3. Algorithms and Complexity
4. Programming Languages
5. Architecture and Organization
6. Operating Systems
7. Net-Centric Computing
8. Human-Computer Interaction
9. Graphics and Visual Computing
10. Intelligent Systems
11. Information Management
12. Software Engineering
13. Social and Professional Issues
14. Computational Science and Numerical Methods

Dijkstra said that to call the field "Computer Science" is like calling surgery "Knife Science". He noted that departments of Computer Science are exposed to a permanent pressure to overemphasize the "Computer" and to underemphasize the "Science."

This tendency matches the inclination to appreciate the significance of computers solely in their capacity of tools.

Computer Science does not deal merely with computer use, technology or software. It is a science that encompasses abstract mathematical thinking and includes an element of engineering. The mathematical element is expressed in finding solutions to problems, or in proving that the solutions do not exist, while the engineering element demands skills for designing complex software systems.

6. SCIENTIFIC METHODS OF CS

Basically in as far as CS is a science we find all features of classical scientific methods in it. Our scheme of Figure 2 is applicable here as well.

What is specific for CS is that its object of investigation is an artifact (computer) that changes concurrently with the development of theories describing it and simultaneously with the growing practical experience in its usage. Computer in 1940s is not the same as computer in 1970s that is different from computer in 2002. Even the task of defining what is computer year 2002 is far from trivial.

6.1 Theoretical Computer Science

Concerning Theoretical Computer Science, which adhere to the traditions of Logic and Mathematics, we can conclude that it follows the very classical methodology of building theories as logical systems with stringent definitions of objects (axioms) and operations (rules) for deriving/proving theorems.

The key recurring concepts fundamental for computing are [10]:

– Conceptual and formal models
– Levels of abstraction
– Efficiency

Data models [10] are used to formulate different mathematical concepts. In CS a data model has two aspects: the values that data objects can assume and the operations on the data. Here are some typical data models:

– *The tree data model* (the abstraction that models hierarchical data structure)
– The list data models (can be viewed as special case of tree, but with some additional operations like push and pop. Character strings are an important kinds of lists)
– The set data model (the most fundamental data model of Mathematics. Every concept in Mathematics, from trees to real numbers can be expressed as a special kind of set)
– The relational data model (the organization of data into collections of two-dimensional tables)
– The graph data model (a generalization of the tree data model: directed, undirected, and labeled)
– Patterns, automata and regular expressions
  A pattern is a set of objects with some recognizable property. The automaton is a graph-based way of specifying patterns. Regular expression is algebra for describing the same kinds of patterns that can be described by automata.

Some of the central methodological themes in theoretical Computer Science (inherited from Mathematics) are *iteration, induction and recursion*.

Iteration. The simplest way to perform a sequence of operations repeatedly is to use an iterative construct such as *for-* or *while-*statement.

Recursion. Recursive procedures call themselves either directly or indirectly. This is self-definition, in which a concept is defined in terms of it self. (E.g. a list can be defined as being empty list or as being an element followed by a list). There is no circularity involved in properly used self-definition, because the self-defined subparts are always “smaller” than the object being defined. Further, after a finite number of steps, we arrive at the basis case at which the self-definition ends.

Induction. Inductive definitions and proofs use basis and inductive step to encompass all possible cases.

In short: theoretical computer science seeks largely to understand the limits on computation and the power of computational paradigms. Theoreticians also develop general approaches to problem solving.

One of theoretical computer science's most important functions is the distillation of knowledge acquired through conceptualization, modeling and analysis. Knowledge is accumulating so rapidly that it must be collected, condensed and structured in order to get useful.

6.2 Experimental Computer Science

Experimental computer science is most effective on problems that require complex software solutions such as the creation of software development environments, the organization of data that is not tabular, or the construction of tools to solve constrained optimization problems. The approach is largely to identify concepts that facilitate solutions to a problem and then evaluate the solutions through construction of prototype systems [11].

Experiment in different fields (search, automatic theorem proving, planning, NP-complete problems, natural language, vision, games, neural nets/connectionism, machine learning) is also used in CS, and is described by methodology of Figure 2.

6.3 Computer Simulation

In recent years, computation, which comprises computer-based modeling and simulation, has become the third research methodology, complementing theory and experiment.

Today, computing environments and methods for using them have become powerful enough to tackle problems of great complexity.

Mastery of Computational Science tools, such as 3D visualization and computer simulation, efficient handling of large data sets, ability to access a variety of distributed resources and collaborate with other experts over the Internet, etc. are now expected of university graduates, not necessarily Computer Science majors. Those skills are becoming a part of scientific culture.

With the dramatic changes in computing, the need for dynamic and flexible Computational Science becomes ever more obvious.

Computational Science has emerged, at the intersection of Computer Science, applied Mathematics, and science disciplines in both theoretical investigation and experimentation.

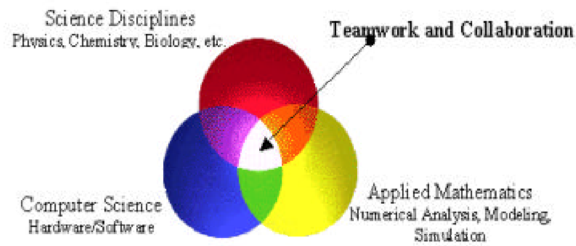

Figure 4. Computational Science emerges at the intersection of Computer Science, applied Mathematics and science disciplines.

Computer simulation makes it possible to investigate regimes that are beyond current experimental capabilities and to study phenomena that cannot be replicated in laboratories, such as the evolution of the universe. In the realm of science, computer simulations are guided by theory as well as experimental results, while the computational results often suggest new experiments and theoretical models.

In engineering, many more design options can be explored through computer models than by building physical ones, usually at a small fraction of the cost and elapsed time.

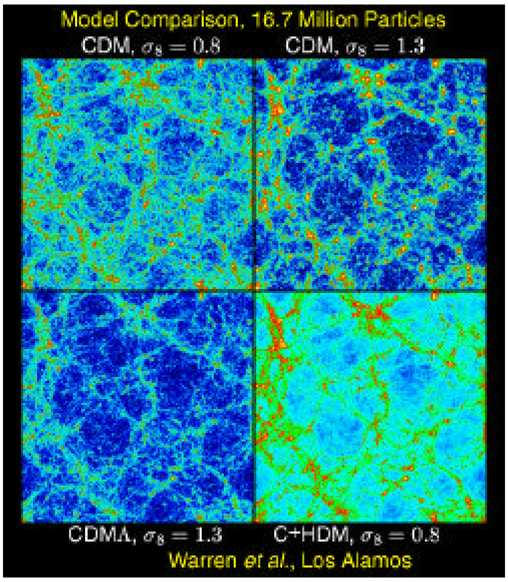

Figure 5 Simulation: Comparison of four astrophysical *N*-body Cold Dark Matter Model variants.

Simulations such as these galaxy formation studies can only be conducted on very powerful computers.

Science often proceeds with bursts of intense research activity. Even though the term "simulation" is old, it reflects the way in which a good deal of science will be done in the next century. Scientists will perform computer experiments in addition to testing scientific hypotheses by performing experiments on actual physical objects of investigation.

One can also say that simulation represents a fundamental discipline in its own right regardless of the specific application. If Computer Science has its basis in computability theory, then Computational Science has as its basis computer simulation.

Let's take some of the key focus areas of the past to shed light on the potential or existing role that simulation plays in each of them:

Chaos and Complex Systems: The idea that one can observe complexity within a structurally simple deterministic model is of fundamental interest. Qualitative topological phase space features of linear systems may be determined statically but simulation must be used for nonlinear systems.

Virtual Reality: Virtual reality is to *immerse* the analyst within the simulated world. Although, it is often seen as being synonymous with man-machine hardware interfaces, the technology must incorporate methods for building dynamic digital (virtual) worlds, which is a typical problem of computer simulation.

Artificial Life: Artificial life is an outcome of Computational Science that challenges our definition of the term *experiment.* An experiment in artificial life is one where a computer program is written to simulate artificial life forms, often carrying along metaphors such as genetic reproduction and mutation.

Physically Based Modeling and Computer Animation: Within computer graphics, there has been a noticeable move forward in the direction of physically based modeling (constraint-based models derived from physical laws).

7. CONCLUSIONS

In spite of all characteristics that differ the young field of Computer Science from several thousand years old sciences such as Mathematics and Logic, we can draw a conclusion that Computer Science contains a critical mass of scientific features to qualify as science.

From the principal point of view it is important to conclude that all modern sciences are very much influenced by technology. That is a natural consequence of the fact that the research leading to the development of modern sciences is very tightly bound to technology. This is very much the case for Biology, Chemistry and Physics, and even more the case for Computer Science that is clearly influenced by industry via engineering.

Engineering parts in the Computer Science often have connection to the hardware aspects of computer, but they even appear in form of software engineering.

Theoretical Computer Science, on the other hand, is scientific in the same sense as theoretical parts of any other science. It is based on solid ground of Logic and Mathematics.

The important difference is that the computer (the physical object that is directly related to the theory) is not a focus of investigation (not even it he sense of being the cause of certain algorithm proceeding in certain way) but it is rather theory materialized, a tool always capable of changing in order to accommodate even more powerful theoretical concepts.

8. REFERENCES

[1] http://www.chem.ualberta.ca/~plambeck/udc/
Universal Decimal Classification (UDC)

[2] http://www.tnrdlib.bc.ca/dewey.html
Dewey Decimal Classification

[3] http://www.computer.org/education/cc2001/index.htm
Computing Curricula 2001
Denning, P.J. et al. Computing: as a Discipline. Commun. ACM 32, 1 (January 1989), 9-23.

[4] Popper, K.R. The Logic of Scientific Discovery, NY: Routledge, 1999

[5] Carnap, R. An Introduction to The Philosophy of Science, NY: Basic Books, 1994

[6] Kuhn, T. The Structure of Scientific Revolutions, Chicago: Univ. Chicago Press, 1962

[7] Chalmers, A. What is This Thing Called Science?, Hackett Publishing Co., 1990

[8] Feyerabend, P. Against Method, London, U.K.: Verso, 2000

[9] Peter Wegner, Research Paradigms In Computer Science, Proc. 2nd Int. Conference on Software Engineering, 1976, San Francisco, California

[10] A. V. Aho, J. D. Ullman, Foundations of Computer Science, W.H. Freeman, New York, 1992.

[11] http://books.nap.edu/html/acesc/ Academic Careers for Experimental Computer Scientists and Engineers, National Research Council Washington, D.C.