

An Education

for the

Software Engineer

C. A. R. Hoare

August 1982

Summary.

The main intellectual tool in the mastery of complexity is abstraction, which can be applied at many levels. In Computer Science there are clearly defined levels: electronic circuits, logic design, computer architecture, operating systems, programs, programming in languages at low, intermediate, and high level, program specification and design, and the theory of computation. The concepts within each single level of abstraction can be understood by the amateur; the professional must understand them all, and also the relations between them. He should also be familiar with techniques application areas - numerical techniques, data processing, communications, control, etc. These are topics in the education of the Computer Scientist

The professional engineer, if he is to be more than a technician, must have a deep understanding of all conceivably) 2

of practical and commercial value. This requires development of technical skills, and more — insight into human and social needs, inventive ingenuity, critical faculty, organisational ability, clarity of exposition, persuasive power, and general managerial skills. His success will depend on the use of abstraction at all levels to control the complexity of his designs and ensure the reliability of their implementation. So this must be the overriding objective of an education for a software engineer.

relevant branches of Science. But then he must also be able to apply his understanding to the design and manufacture of products

Levels of Abstraction.

The world's a complicated place:

the random scattering of stars in the sky, and their regular motion ^{round the poles;} the random wandering of the moon and planets and the regular alternation of night and day; the random fluctuations of the weather and the regular succession of the seasons; the occasional stroke of good fortune and the more frequent affliction of famines, pestilence, and war — all these phenomena require convincing explanations.

From the earliest times such explanation has been given in widely comprehensible terms; they were supposed to commemorate the adventures, the pranks, the passions of the gods in their dealings with each other and with the mortals who fall under their power.

(mythological)
These explanations had the advantage of great simplicity and uniformity of concept, of wide applicability and emotional appeal, which have rarely been matched by more modern scientific theories. Perhaps the most successful has been the Newtonian theory of gravitation; which not only explains the celestial phenomena of such great concern to early man, but also to matters of greater concern to modern man — such as the bouncing of balls and the aiming of bombs. Materialist philosophers have speculated that if at some moment the position and momentum of every particle in the universe were known, then the entire future of the universe could be predicted.

With the further progress of natural science, we have come to realise that the world is much more complicated than was ever conceived by earlier mythologists or philosophers. We no longer expect that all phenomena can be successfully explained within a single conceptual framework or by a single physical equation. Each question must be formulated and answered in the terminology of the appropriate branch of science. Famine may be explained and perhaps avoided by the study of biology; pestilence by the study of medicine, and the distribution of stars by the study of cosmology; ~~the~~ ^{the} irrationality of war ~~seems~~ still ~~to~~ escapes explanation by the methods of scientific reason.

Perhaps one day all branches of science will be unified under some single abstract theory, as comprehensive as that of materialist philosophy, as attractive as that of poets and mythologists. To find a theory which forges a link between previously isolated areas of knowledge has been the driving ambition

and greatest achievements of modern science.

Thus Molecular biology explores the area between chemistry and medicine, cosmology provides a bridge between astronomy and physics.

This Newtonian mechanics was a unification of theories of terrestrial and celestial motion.

But the discovery of such links in no way invalidates or supercedes the theories which have been unified. A successful scientific theory is a selection, ^{approximation}, and abstraction from the phenomena of the natural world. It provides a conceptual framework within which these phenomena can be described, summarised; and it formulates a mathematical calculus by which they may be predicted and controlled. The validity and usefulness of the techniques and formulae of a simple scientific theory are only increased if it is shown, by more complicated reasoning and experimentation, to be a simplification or approximation to some more general underlying theory.

Thus the practical advice of the baseball coach is supported, not superseded by Newtonian laws of motion; the practical formulae of the gunnery officer are not invalidated by the theory of relativity, the breeding programmes of the scientific agriculturist are not immediately affected by the latest discoveries of molecular genetics.

9

In summary, we see the ^(ideal) structure of modern science not as an universal all-embracing theory but as a range of specialised abstractions, each applied to a limited range of phenomena.

These abstractions may be arranged in a multilevel hierarchy, in which each of them can be explained and justified in terms of the abstraction on the level below. The concepts and formulae of each separate theory may be simple enough to be understood by the interested layman, and applied by the competent technician. But the reasoning and constructions which show how one level of abstraction is based upon another may be much more

complicated: they are the preserve of the professional scientist.

Among the natural sciences, the one that exhibits most clearly the levels of abstraction is Physics. In general, each level deals with particles of a different size, and with processes manifesting themselves over different timescales (see table 1). The importance of these

levels of abstraction is not that each can't be explained in terms of the one below, but rather that in practice there are good reasons for avoiding such a mixture of levels.

The mechanics of solid bodies are best understood without detailed consideration of their molecular structure. It is a pleasing paradox that some of the most promising advances in modern physics are in the relations between

the highest and lowest levels, between cosmology and quarks

	size	timescale
Quarks		?
Subatomic Particles	10^{-}	
Atoms	10^{-} m	
Molecules and crystals	10^{-}	
Mechanics	—	seconds
Astronomy		
Cosmology		

Fig 1. Levels of abstraction in modern physics.

Levels of abstraction in modern physics.

Another subject in which levels of abstraction are can be discerned is Pure Mathematics. Here again, each level can be fruitfully studied and used in relative isolation from the other levels; although an understanding of the

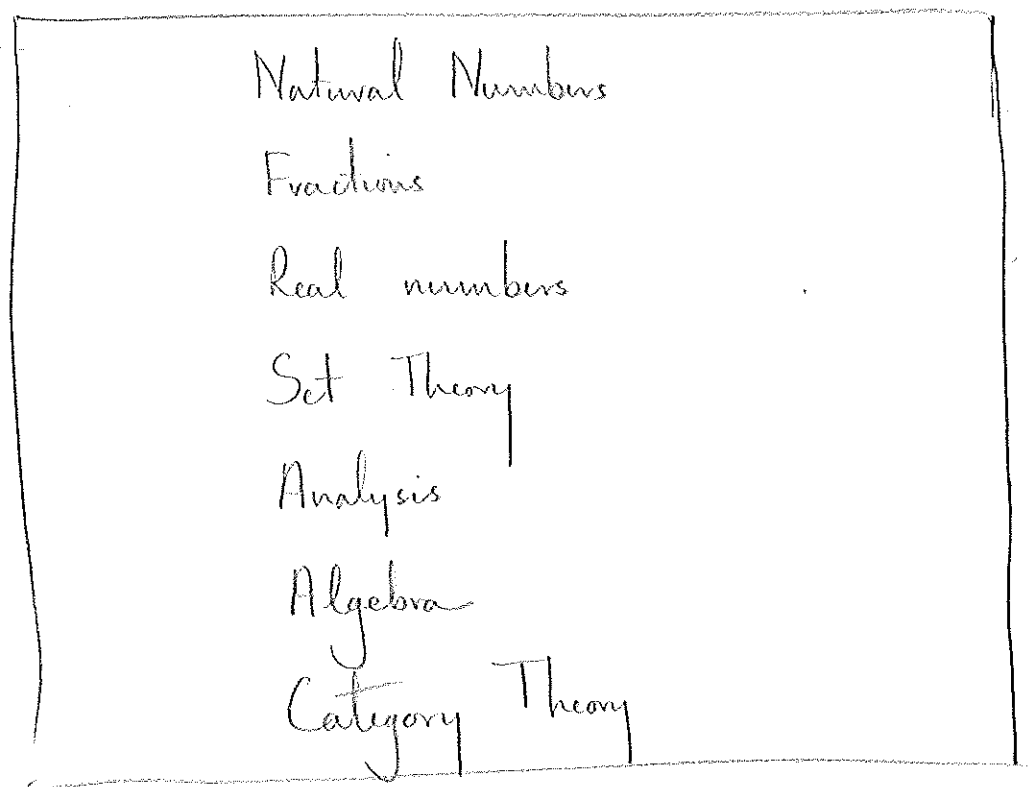


Fig 2. Levels of abstraction in Mathematics

relationship between the levels is most illuminating for the serious mathematician.

2. Computer Science

Stored-program digital computers are the most complicated artifacts devised by human ingenuity. To design them, to understand them and to use them effectively, we need the same powerful intellectual tools which have been successful in mastery of the complexities of the physical world or of mathematics. So it is not surprising to find in Computer Science a clearly defined series of levels of abstraction (Fig 3)

electronic circuits
digital logic
computer design
microprogramming
assembly code
programming languages
algorithms.
data structures
programming methodology
theory of computation.

Fig 3. levels of abstraction in computer science.

Each successful level of abstraction must be to some extent self-contained and self-sufficient; it can be learned and understood and used to advantage without reference to the layers above and below. Thus a logic designer should not exercise his knowledge of electronics or of microprogramming; and the microprogrammer should ignore what he knows about logic design and assembly code. Indeed, if required, these tasks could be delegated to technicians, who know nothing about underlying theory, or the relationship between related tasks, or the contribution which they are making to the whole design.

But this degree of specialisation is not appropriate for the scientist or the engineer, for the professional or for the manager. Of course, he will be willing and able to develop specialised skills, and exercise them for sustained periods. But he will not be satisfied with that: he also wants to understand the full range of related tasks, and the relationships between them. And that is the proper purpose of a professional education, as opposed to the more specialised training of the technician.

As a result of the relative isolation of the levels of abstraction, they form an excellent structure for planning the education of the computer scientist. Each level forms the subject matter of a single examinable course. Each course includes the relevant aspects of formal mathematical theory and calculus; each course is supported by practical work to inculcate the relevant skills. Whenever possible, each level should be proved to be firmly based on the level below, and to provide a secure basis for the level above.

In some cases, the links between the levels are too complicated to be studied as part of either of the related levels. For example, the link between assembly code programming and high level languages is provided by the compiler for the language. The study of compiling methods is worthy of a substantial separate course. Such a course is important not only because of its intrinsic interest as a bridging subject, but because it illustrates many of the most important considerations in the design of computer systems programs:

- (1) Clear understanding of the subject matter of the program, i.e., the language being compiled
- (2) Techniques for the analysis and processing of symbolic text
- (3) Design of a human interface with good recovery from human error.
- (4) Clear structure and documentation needed for any large program.

Similar benefits can be obtained from the study of operating systems, or indeed any significant application area of computing, for example commercial data processing numerical methods in science and engineering communications protocols, coding, etc.

Physical Science

quarks and quantum chromodynamics

elementary particles

atoms

molecules

crystals

materials

mechanics

astronomy

cosmology

77

10⁴⁶ sec

Engineering

Market insight - human driven
invention - technology driven
evaluation
design
prototyping
production engineering
toolbuilding
quality control
sales

planning
budgeting
administration
management

computer science

electronics circuits

logic design

digital logic

computer architecture

computer design,

assembly programs

assembly coding

high level languages

algorithms

data structures

human w operating systems

specifications

man/machines
applications - interface

theory of computation.

standards.

communications

