The Computer Revolution In Philosophy (1978)

Aaron Sloman

Book contents page

This chapter is also available in PDF format here.

CHAPTER 10

MORE ON A.I. AND PHILOSOPHICAL PROBLEMS

10.1. Introduction

Chapter 3 included a long list of philosophical questions of the form ’How is X possible?’ Patient readers will find many points of contact between those questions and the topics of the last few chapters, especially the sort of work in computer vision described in Chapter 9. In this chapter I shall comment further, in a necessarily sketchy, shallow and speculative fashion, on some of the connections between philosophy and recent steps towards the design of a mind. Much of the discussion is speculative because I shall be talking about types of computing systems whose complexity exceeds anything so far programmed. But work already done in A.I. points clearly in the directions I assume to be feasible.

Not all the philosophical problems I shall be referring to are of the form ’How is X possible?’ But the first one is: namely how is it possible for there to be a distinction between conscious and unconscious mental processes’ Alternatively, how is it possible for some, but not all, of the contents of our minds to enter into our conscious experience? This topic will be discussed at some length, after which a collection of loosely related problems will be touched on.

10.2. Problems about the nature of experience and consciousness

’What is consciousness?’ is a very tricky question, for several reasons. A full analysis of what we ordinarily understand by the words ’conscious’, ’consciousness’, and related expressions, such as ’awareness’, ’self consciousness’, ’experience’, and the like, would show that they are very complex and subtle. Such an analysis, using the sorts of techniques outlined in chapter four, should, ideally precede an attempt to provide some sort of scientific explanation of phenomena involving consciousness.

I shall not go into such a detailed analysis now. But I want to say something -- not about the most general sense of the word ’conscious’, which includes usages like ’I’ve been conscious for several months that I am likely to lose my job soon’, which refers to some knowledge or belief -- but about the kind of distinction we make between things that we are currently conscious of and things we are not, especially things in our own minds. I want to try to relate this distinction to some computational considerations.

It is obvious that besides conscious mental processes there are unconscious or subconscious ones, such as the decisions about gear changes, steering and so on taken by an experienced car driver, the recognition of syntactic structure in understanding spoken and written language, and the detailed analysis and interpretation processes involved in perceiving a complex scene or picture. (Chapter 9.) Moreover, what a learner is painfully conscious of may later be handled unconsciously -- like gear-changing while driving a car, or using grammatical constructs in a second language. So there need be no difference in the *content* of conscious and unconscious processes.

Although it is obvious that there is a difference, it is very difficult to analyse this difference between what we are and are not conscious of, especially as there are so many borderline cases -- like finding something odd without being aware of what is odd about it. Were you previously *conscious* of the fact that you were reading print arranged in horizontal lines or was it unconsciously taken for granted? How is this different from being conscious of the lines of print? Is a sleep-walker who clearly opens a door in order to go through, conscious of the door and aware that it is shut? Is he conscious that he is opening it? While reading a gripping story you may be very conscious of what is going on in the story, but hardly aware of what is on the page. A good quick reader is conscious of some of what is on the page, but not necessarily all the letters composing words he reads. And he may be too engrossed in what he is reading to be conscious of *the fact* that he is reading.

For the past few minutes you have probably been conscious of the fact that you were reading, but were you also conscious of being conscious of it? And were you conscious of that too? How far are you prepared to go in saying that you are conscious of being conscious of being conscious of . . . etc.?

That was merely a reminder that what may at first seem to be a clear and obvious distinction is often very slippery when looked at closely a typical philosopher’s delight! Do not be misled by rhetorical invitations to grasp the essence of consciousness, or experience, or mind, by examining your own current awareness. Introspection is not as easy or informative as some think!

But there is a distinction, however slippery it may be. So we can ask questions like: what is it for? How does it come about that we are conscious of some of our mental states and processes, but not others? What is special about the former? Would we have any need to build in such a distinction if we were designing a person, or an intelligent robot? What are the preconditions for such a distinction to arise in a complex information-processing system?

If, as suggested in chapter 6, we can make a distinction between relatively central administrative processes and the rest, then perhaps we can use this as a basis for analysing the distinction between what the system is conscious of and what it is not, roughly as follows:

> What the system is currently conscious of includes all the information available to the central decision-making processes, whether or not decisions are actually influenced as a result. The system would be *self-conscious* to the extent that the information available to these processes included information about the system itself e.g. information about its location, its current actions, its unfulfilled purposes, or even about what it is currently conscious of! (Compare Minsky’s ’Matter Mind and Models’.)

Let us try to clarify this a little, recapitulating some points from chapter 6. The central processes are those which, among other things:

a) choose between different motives, and control major processes of deliberation and planning, like forming new overall long-term aims and strategies,

b) assign tasks and allocate resources to sub-processes,

c) resolve conflicts between different sub-processes (for instance if the desire for water generates the intention to go in one direction whereas the desire to avoid the tiger near the water-hole generates a desire to go in the opposite direction)

d) set up monitors to watch out for occurrences which might be specially important in relation to current activities,

e) decide what to do about new information from high-level monitors,

f) in some systems they might also control the organisation and cataloguing of major information stores used by many different kinds of sub-processes.

There would not be so great a need for any such centralised process if there were not the possibility of conflicts. The body cannot be in two places at once, the eyes cannot look in two opposed directions at once, and there are limited computational resources, so that expensive processes cannot all run simultaneously (e.g. if one of the main information work-spaces has a small capacity). There might also be conflicts of a more subtle sort, for example conflicts between different ways of interpreting some information which is not at present relevant to any on-going activity, but which *might* be. In all these cases, sub-processes will generate conflicting goals, plans and strategies, and so there must be some means of resolving the conflict, taking into account the needs of the whole system (the need to avoid serious injury, the need for food, the need for well-organised catalogues and information stores, the need to go on collecting information which might be useful sometime, the need to develop new abilities and improve old ones, and so on).

The need for global decision-making processes would be further reduced if the system were less flexible, that is, if it were not possible to change the nature and aims of different sub-processes. Where a complex system has a relatively fixed structure, there will be no need for decisions about what the structure should be!

What I have called ’central’ processes need not be located centrally in a physical sense: indeed, for reasons given in chapter 5 and elsewhere, they need not have *any* specific physical location. For example, in a nation where all citizens vote on every major policy decision, everybody is part of the central process.

Further, the central processes need not all be under the control of some single program: the central administrator may itself be simply a collection of sub-processes using certain stores of information, but changing in character and strategy from time to time, like the political party in power. Its function in the total system is what defines the central process, or collection of processes.

If lots of separate sub-systems could happily co-exist without any conflicts, and without any need for or possibility of a co-ordinated division of labour, then there would not be a role for any kind of centralised decision-making. Alison Sloman informs me that there are several kinds of organisms which live together in co-operative colonies, but which do not need the sort of global decision-making I am talking about. Coral is an example. If, like most plants, such a colony cannot move or has no control over its movements, or if which way it moves does not matter, then there cannot be conflicts about which way it should move. If a system does not have eyes, then there cannot be conflicts or decisions about which way it should look. This suggests that the evolution of organisms with a distinction between conscious and unconscious processes may be closely related to the evolution of forms of symbiosis and co-operation in complex tasks, and the differentiation of functions.

(This line of thought also suggests that it may be possible to make a distinction between what a human social system is and is not conscious of, if it is a relatively integrated system. Of course, we must not expect the distinction to be any less blurred and slippery than it is when applied to individual people.)

So, perhaps the distinction between what we are and are not conscious of at a particular time, is concerned with the difference between information which is made available to. or used by, central administrative processes, and information which is not. There will be many processes which continue without any notice being taken of them by the central administrator, and at each moment there is an enormous amount of unused information present in stores of various kinds. There is no point cluttering up the central decision-making with all the details of the sub-processes: the task of relating all the information would be too unmanageable. So censorship of a sort is a prerequisite for normal functioning of such a system, rather than an oddity to be explained. (This principle is integral to the design of the POPEYE program described in chapter 9.)

> **[[Note added Sept 2001:**
> In the years following publication of this book many researchers have attempted to avoid the need for any kind of central administrative mechanism by postulating networks of cooperative and competing mechanisms through which global decisions and behaviour can emerge. Typically this requires the notion of some sort of common *currency* in terms of which the relative importance of different needs and goals and plans can be evaluated by local comparisons, and possibly some sort of voting scheme for combining the preferences of different components of the system.
>
> Despite the popularity of such ideas I suspect they are appropriate only to problems where there is no possibility of a well structured solution based on a clear understanding of the different sub-goals, their relationships, the options for action, the possibilities for compromise or for optimal sequencing. Where attempts are made to base decision-making entirely on numerical computations, e.g. using probabilities and utilities, it often turns out (in AI and in government procedures) that reliance only on numerical processes loses much information, by comparison with descriptive methods. A consequence is that good solutions cannot be found except in simple cases.
>
> The idea of a high level unitary decision-making process for resolving conflicts on the basis of a global viewpoint is often re-invented. E.g. See P.N. Johnson-Laird, *The Computer and the Mind: An Introduction to Cognitive Science*, 1993 (2nd Edition). He draws an unfortunate analogy with operating systems, unfortunate (a) because typically operating systems are concerned with huge amounts of low level management in addition to the more central global decision making, and (b) because an operating system can often become subservient to a more intelligent program running within the operating system, e.g. AI programs controlling a robot. **]]**

It is possible for perceptual sub-processes which do not influence the central processes at all, to produce modifications of the store of beliefs, and help to control the execution of other sub-processes. They may even influence the *central* processes at some later stage -- a possibility taken for granted by advertisers and propagandists. This amounts to a form of unconscious perception, differing from conscious perception only in its relationship to the central processes. So from the present viewpoint, the existence of unconscious mental processes is in no way puzzling.

We can *become* conscious of some, but not all, of the things in our minds of which we are not conscious. Much of the information which is not accessed by central processes *could be* if required. There are all sorts of things in your memory, of which you are currently not conscious (though if asked you might say you have been aware of them for several years!), but which you could become conscious of if you needed the information.

The same is true of much of the information processed by our senses: you may become conscious of the humming noise in the background which you previously did not notice, because someone draws your attention to it, or because it stops, or even because you simply decide to listen to your surroundings. However, some things are not accessible. Why not?

There are several different sorts of reasons why information about a complex system may be inaccessible to the central processes. Here are some, which might not occur to someone not familiar with programming.

a) As already pointed out, many sub-processes will acquire, use, or store information without any need to notify central processes. They will use their own, private work-spaces. If the information is not recorded in globally accessible records. there may be no way the central processes can get at the information, for instance information in peripheral perceptual processes. The sub-processes may be incapable of being modified so as to make them store information elsewhere, and it may not be possible to give monitors access to their ’innards’. This is especially likely to be the case if the mechanisms are ’hard-wired’ rather than programmed. In a computer it is relatively (!) easy to change the behaviour of programs, whereas changing the behaviour of the underlying physical machine may be impossible.

b) When a program is executed in a computer, it may keep records of some of its activities in examinable structures, but not all of its activities. The records enable a computer to answer questions about what it is doing or has done, and, more importantly, enable it to do things more intelligently, since different sub-tasks can be explicitly compared with one another, so that learning and self-control can occur. Storing explicit records of processes takes up space and extra processing power, but it may provide much greater flexibility, including the ability to learn from mistakes. This is what happens in Sussman’s Hacker system when it executes ’in careful mode’ the programs it has designed (Sussman, 1975). So one source of inaccessibility may be simply the fact that although a program does things, it does not keep any records which may be examined later, even a very short time later.

c) Even if information is present in some store, it may not be accessible until suitable entries have been made in indexes or catalogues. So some facts about what is going on in our minds may be recorded quite explicitly, yet never indexed properly. This could prevent central processes ever finding out about them.

d) Whether information actually present is accessible or not can also depend on peculiarities of the processes which attempt to access them. Some processes may have a built in assumption that all information relevant to them can be found via a particular sub-catalogue. (Like people who think that the only good books on philosophy are to be found in the philosophy section of the library.) Hence information may be inaccessible at certain times simply because the searching is done in too inflexible a fashion. Suitable forms of learning may improve the flexibility of our information-accessing processes, making us more conscious of what we are doing. (However, there will usually be a price to pay for increased flexibility such as reduced speed: another trade-off.)

e) Among the events which are not recorded explicitly, some, but not all, can be readily recomputed from items which are recorded. So, in some cases, inaccessibility may be accounted for in terms of inadequate records being kept, or inadequate inference procedures for reconstructing what happened from available records.

f) Some of the explicit records of what is going on may be inaccessible because the need to refer to them has not been recognised by the central process. Perhaps it failed to set up appropriate monitors (chapter 6) because of poor procedures for the task in hand. For example, when learning to play a musical instrument people often find it very hard to learn to keep on listening to important aspects of their own performance which they need to hear to control their playing, even though they have no difficulty in listening to someone else. Similarly, many teachers fail to attend to the evident effects of their behaviour on their pupils.

g) People often react to cues by jumping to conclusions about something, and thereafter fail to examine the readily available evidence further to check whether the conclusion is correct. Single-minded or simple-minded programs may behave in the same way. And very often this is a very sensible way to behave, if rapid decisions have to be taken (see chapter 9). However, if the strategy is firmly embedded in a collection of procedures for interpreting certain information, then some aspects of the information may never be examined properly.

h) The system may lack the descriptive and interpretative abilities required for perceiving the significant relationships between items of information which are readily accessible. (Compare chapter 9, and remarks about concepts in chapter 2.) Suitable concepts, and training in their use, may be required before important facts can be noted. If you have never grasped the concept of symmetry you cannot be conscious of the symmetry in a pattern. Someone who has not learnt to think about the difference between valid and invalid arguments cannot be aware of the validity or invalidity of an argument. A child who has not learnt to think about grammatical categories cannot be aware that he is, or is not, matching the number of a verb and its subject.

(I believe that much of what Marxists refer to as ’false consciousness’, like the inability of people to see themselves as exploited, can be accounted for in terms of a lack of some of the analytical and interpretative concepts required. What needs to be explained, then, is not why people are not conscious of such facts, but how it is possible for them ever to learn the concepts which can make them conscious.)

i) Some processes may use a temporary work-space which is not fully integrated with the enduring memory structures, but instead gets re-used frequently. While information is in this temporary store it may be as accessible as anything else -- but if it is not accessed before the space is re-used it will be permanently lost. So the reason the records are inaccessible to the central processes may be that searches are always carried out too late.

Much of what we do may involve such rapidly re-used storage so that if asked about details shortly after doing things we cannot recall exactly what happened. Perhaps the activities of a sleep-walker who seems to be fully conscious while walking about also use such temporary storage space for records which would normally be linked to more enduring structures. (None of this presupposes that there is any physical difference between the permanent and the temporary storage locations, nor in the mechanisms for accessing them. It may even be possible for ’permanent’ records to be obliterated and the space re-claimed for temporary storage! A lot depends on the storage medium, about which very little is known in the case of humans.)

What I have been driving at is that what is hardest to explain is not why some things are inaccessible, but how things ever become accessible to central processes. We do not need to postulate mechanisms for preventing things becoming conscious: mere *lack* of a mechanism, or activity, may explain that. However there may be explicit suppression or censorship too.

We have already seen that there is good reason for arranging that only a subset of all goings-on be reported centrally. So sub-processes may have explicit instructions about what to report and what not to report. Moreover, it is necessary for these instructions to be modifiable in the light of current needs and expectations. So the central administrator may have some control over what gets reported to it. Thus there is plenty of scope for it to give explicit instructions preventing certain categories of information being recorded, or reported to globally accessible stores.

So some items may be inaccessible as a direct result of policy decisions within the system (as Freud suggested). Records of these policy decisions may themselves be inaccessible! (Many of these points will be quite obvious to administrators, both corrupt and honest.) Further study of this topic should illuminate various sorts of human phenomena, desirable and undesirable.

I have already warned against the assumption that there is necessarily a *unique* continuing process with the centralised decision-making role. There might be a number of relatively self-contained sub-processes which gain control at different times. If they each have separate memory stores (as well as having access to some shared memory), then we can expect schizophrenic behaviour from the system. Perhaps this is the normal state of a human being, so that, for example, different kinds of central processes, with different skills, are in control during sleeping and waking, or in different social settings.

Maybe only a subset of what constitutes a central administrator changes during such switches, for instance, a subset of the motivational store and a subset of the factual and procedural memory. Then personality has only partial continuity.

It is possible (as I believe Leibniz claimed) that instead of there being one division between what is and is not conscious in a complex system, there may be many divisions one for the system as a whole, and more for various sub-systems. If there is something in the argument about the need for some centralised decision-making in the system as a whole, then the same argument can be used for the more complex sub-systems: considered as an organic whole, there may be some things a sub-system can be said to be conscious of, and others which it cannot.

This would be clearest in a computer which controlled a whole lot of robot-bodies with which it communicated by radio. For each individual robot, there might be a fairly well-integrated sub-system, aware of where the robot is, what is going on around it, exactly what it is doing, and so on. Within it there will be sub-processes and information-stores of which it is not conscious, for the reasons already given (and no doubt others). Similarly within the total system, composed of many robots, there will be some kind of centralised process which is not concerned with all the fiddly details of each robot, but which knows roughly where each one is, knows which tasks it is performing, and so on. It may be capable of attending closely to the things an individual robot is looking at, thinking about, feeling, etc., with or without its knowledge, but will not do this all the time for all of them. So individual robots may be aware of things the system as a whole cannot be said to be aware of, and vice versa. Worse, the whole thing might itself be only a part of a still more complex yet centrally controlled system!

Maybe that is the best way to think of a person: but if so we shall not fully understand why until our attempts to design a working person have forced such organisations on us.

We need further analysis of the sorts of computational problems which might lead to subdivisions of administrative functions, and the reasons why the development of individual systems might go wrong, leading to too many relatively independent sub-systems, or to too little communication or shared structure between them. Psychiatry and education might hope to gain a great deal from such studies. Perhaps the same is true also of political science.

We are at present nowhere near an adequate analysis of the concept of conscious experience, and related concepts. But it seems that in investigating the different forms of self-awareness required by intelligent mechanisms we have a far better chance of getting new insights than from the typical style of philosophical discussion on this topic, which all too often is a mixture of dubious introspective reports and dualist or anti-dualist prejudice.

10.3. Problems about the relationships between experience and behaviour

In the course of analysing and interpreting a complex image a computer may generate a very large number of sub-processes, and build up many intricately interrelated symbolic structures. (See Chapter 6, Chapter 8 and Chapter 9.) Although these processes and structures are used temporarily in subsequent analyses, the organisation of the system may make it quite impossible for the program to express in its *output* anything more than a brief summary of the end product, for example, ’I see a man, sitting at a table covered with books and papers.’ There may be several different reasons for such restrictions.

For instance the available output *medium* may be ill-suited to represent the rich detail of the internal structures (as a linear string of words is ill-suited to represent a complex map-like network). Or the processes and structures may be set up in such a way that output mechanisms cannot access them, for any of the sorts of reasons mentioned in discussing consciousness.

So crude behaviourist analyses of statements about the detailed experiences of the computer must be rejected. Experience, conscious and unconscious, in humans, animals and machines, may be much richer than anything their behaviour can reveal.

But even more subtle dispositional or behaviourist analyses (in terms of how the behaviour *would have* been different if the stimuli had been different, e.g. if probing questions had been asked) may be inappropriate for the program need not allow for *any* behavioural indications of some of the fine details of the internal analysis.

For example, a compiler which translates high-level programs into machine code may be written in such a way, that it is impossible (without major re-programming) to obtain a print out of some of the structures temporarily created during the translation process, for instance the temporarily created ’control-structures’. After all, its main function is not to print out records of its own behaviour, but to translate the programs fed into it.

The situation is more complex with an operating system. One of the tasks of an operating system may be to manage the flow of information (inwards or outwards) between sub-processes in the computer and various devices attached to it. If it is required to print out details of how it is managing all the traffic, then this adds to the traffic, thereby changing the process it is attempting to report on. This sort of thing makes it very difficult to check on the workings of an operating system. But the main point for present purposes is that there are computational systems which cannot produce external behaviour indicating features of their internal operation without thereby significantly altering their operation. There is no reason to doubt that this is true of people and animals.

All this means that the scientific study of people and animals has to be very indirect if they are computational systems of the sort I have been discussing. In particular, the lack of any close relation between inner processes and observable behaviour means that theorising has to be largely a matter of guesswork and speculation. The hope that the guesswork can be removed by direct inspection of brains seems doomed. You will not find out much about how a complex compiler or operating system works by examining the ’innards’ of the computer, for they are programs, not physical mechanisms. The only hope of making serious progress in trying to understand such a system is to try to design one with similar abilities.

> **Note added 25 Sep 2009**
> ‘‘The only hope’’ is too strong. Rather I should have written something more like ‘‘The only hope of making serious progress in trying to understand such a system is to combine as many different empirical investigations, of what individuals do, how they develop, how the species evolved, what the brains do, how they do it, with attempts to design machines with similar abilities in order to understand what the problems are that had to be solved, and in order to test partial solutions.

10.4. Problems about the nature of science and scientific theories

Computer models of visual perception are attempts to answer questions of the form ’How is X possible?’ for instance, ’How is it possible to interpret an untidy collection of visual data as representing such and such a scene?’ and ’How is it possible for locally ambiguous image fragments to generate a unique global interpretation?’. So they provide a further illustration of the claim in chapter 2 that science is concerned with discovering and explaining possibilities.

Moreover, although such models are rich in explanatory power, since they can explain some of the fine structure of visual abilities, they do not provided a basis for prediction. This is because, like many explanations of abilities, or possibilities, they do not specify conditions under which they will be invoked, nor do they rule out the possibility of extraneous processes interfering with them. So, how we use our visual abilities (for example, what we notice, how we react to it and how we describe our experiences to others), depends on our desires, interests, hopes, fears, and on our other abilities, rather than merely on what enables us to see. (As Chomsky has often pointed out, competence is not a basis for predicting actual performance.)

An explanatory program will have some limitations. There will be some situations it cannot cope with, for example, pictures which it interprets wrongly or not at all. Predictions of human errors could be based on some of the errors made by the program, and if similarities are discovered, that supports the claim that the program provides a good explanation of the human ability. However, people may use additional resources to cope with the situations where the program goes wrong. For example, some knowledge about the whereabouts of a person may prevent your mistaking another person for her, whereas a program using only visual similarity would go wrong. This ability to recover from mistakes is to be expected if, as explained in chapters 6, 8 and 9, intelligent systems require multiple ongoing processes, some of which monitor the performance of others. So even if it is true that a certain person uses exactly the same strategy as some computer program, in all the cases where the strategy is successful, there need not be a close correspondence between the program’s limitations and the limitations of the person. Explanatory power, then, is not necessarily bound up with predictive power, though it does depend on generative power.

Similar remarks could be made about other sorts of A.I. work. For instance, language-understanding and problem-solving programs are rich in explanatory power in the sense of being capable of generating a variety of detailed behaviours. So they are good candidate explanations of how it is possible for people to behave in those ways. Yet they do not provide a basis for predicting when people will do things. So they do not explain laws.

What this amounts to in computational terms, is that to specify that a collection of procedures and information is available to a system explains capabilities of the system, but does not determine the conditions under which they are invoked or modified by other procedures in the system. So work in computer vision, like much else in A.I. and linguistics research, supports the claim of chapter 2 that explanatory power is related more closely to generative power than to predictive power. Rival explanations of the same abilities may be compared by comparing the variety and intricacy of the problems they can cope with, and the variety of different sorts of behaviour they can produce. When we begin to develop programs which approximate more closely to human competence, we shall have to use additional criteria, including comparisons of implementation details, and of the underlying machines presupposed.

10.5. Problems about the role of prior knowledge in perception

It is possible in principle for a system with little or no initial knowledge somehow to be modified through a long period of interaction with the environment so that it acquires perceptual abilities. However, this sort of learning without presuppositions can only be a relatively blind trial-and-error process. The clearest example seems to be the evolution of mechanisms like perceptual systems in animals. This process of learning with minimal presuppositions apparently requires millions of years and is quite unlike the learning achieved by an individual animal after birth, which is much more rapid and intelligent, especially in humans. So completely *general* theories of learning, not related to knowledge about any specific domain, and capable of explaining only the ability to conduct huge, unguided searches through millions of possibilities, are unlikely to have much relevance to human learning, though they may usefully characterise some evolutionary processes.

I am not claiming that we understand the evolution of intelligent species. In particular, it is not obvious that the blind, trial-and-error learning process continues beyond the earliest stages. A species (or larger biological system) is a complex computational mechanism, with distributed processing power, and as such it may be able, to some extent, to direct its own development just as some species (e.g. humans) already direct the evolution of others (e.g. breeding cattle). (Some people have explicitly recommended generalising that to human evolution.)

As Kant recognised, *intelligent* learning from experience requires considerable prior domain-specific knowledge. Chomsky (1965) makes this point about language-learning, but it is clearly very much more general. This is borne out by attempts to give computers visual abilities. All programs which do anything like perceiving objects and learning about the environment seem to require a rich body of implicit theoretical and practical knowledge. The theoretical knowledge concerns the possible structures of sensory data and the possible forms of ’scenes’ which can give rise to such experience.

The practical knowledge concerns ways of *using* the theoretical knowledge to interpret what is given. Nobody has been able to propose explanations of how an individual might acquire all this knowledge from experience, without prior knowledge to drive the analysis and interpretation of experience.

What we are beginning to learn from such artificial intelligence research is the precise nature of the background knowledge required for various forms of visual perception. For instance, by designing working models we can explore such questions as: what sorts of knowledge about the geometry and topology of images does a visual system require? Which sorts *of general* knowledge about space and *specific* knowledge about particular sorts of objects can enable a rational system to find the best global interpretation of a mass of locally ambiguous evidence without wasting time exploring a host of unsatisfactory possibilities? How much prior knowledge of good methods of storing, indexing, and manipulating information is required?

We also breathe new life into old philosophical and psychological problems about the general categories required for experiences of various sorts, or the sorts of concepts which are grasped by infants. For example, the POPEYE program samples the given image looking for dot-strips unambiguously indicating a portion of a line. If two such fragments are collinear, the program hypothesises that they belong to the same line. Thus it uses the concept *of an object extended in space.* Similarly if a program is to interpret a series of changing images in terms of some sort of continuous experience (as in Weir, 1974, 1977) then it requires the concept *of an object enduring through time,* as Kant pointed out long ago.

These object concepts play an important role in organising and indexing information so that it can be *used.* In order to have integrated perceptual experiences one needs to make use of concepts of objects which in some sense go beyond what is given. The object-concepts are organising wholes with explanatory power. (I am not claiming that these concepts are necessarily used *consciously.* The relationships between this and claims about object concepts made by Piaget and other developmental psychologists remain to be explored. I believe newborn infants are grossly underestimated in this as in other respects.)

When better theories about the presuppositions of different sorts of learning have been developed, we shall be in a much better position to assess the rationality of the processes by which knowledge can be derived from experience.

Philosophers' writings about the relation between knowledge and perception normally ignore all the complexities which come to light if one begins to design a working visual system. In particular, it is usually taken for granted that the contents of our sensory experiences, such as patches of colour, lines, shapes, are somehow simply 'given', whereas work in A.I. suggests that even these are the results of complex processes of analysis and interpretation. So whereas philosophers tend only to discuss the rationality of inferences drawn from what appears to be given, we can now see that there is a need to discuss the rationality of the processes by which what is given emerges into consciousness. I have tried to suggest that this emergence is the result of very complex, usually unconscious, but nevertheless often rational, processes.

10.6. Problems about the nature of mathematical knowledge

As explained in chapter 9, perceptual systems require a great deal of prior (usually implicit) knowledge of the possible structure of their own experiences and possible interpretations thereof. This is what distinguishes a system which analyses or interprets the sensory information it receives, from a device, like a camera, or a tape recorder, which passively records such information.

What the prior knowledge is, and how it should be represented in a usable form, are topics of current research. But it seems to be settled beyond doubt that it includes a certain amount of topology and geometry not all of which can have been acquired from perceptual experience, since it is required for such experience (unless we count the evolution of the human species as experience).

I am not suggesting that children are born with the contents of mathematical text-books in their heads. Much of the knowledge is probably in procedural rather than factual form, and the set of initial concepts is likely to be different from the set of primitives in a mathematical presentation. For example, it is possible that the notion *of straight line* develops only later on, from some kind of more general notion of a line.

We are now faced with the possibility of new detailed explorations into processes by which such a system might become aware of the limits of possible forms of sense-data, the limits of its own interpretation procedures, and the limits on the forms of interpretation it is capable of generating. In this way we may hope to discover new answers to the old question: ’What is the nature of geometric knowledge?’

Already it seems clear that in concentrating on geometry, Kant missed some deeper and more general forms of knowledge concerned with topology, a branch of mathematics which had not been developed at the time. Many other Kantian questions can be reopened in this way, such as questions about the nature of arithmetical knowledge, discussed in chapter 8.

Very little work has been done so far on ways of giving computer programs the ability to discover their own abilities and limitations. The most obvious method is to let a program try all possible combinations of sub-procedures to see what can and cannot be achieved. However, for complex systems this requires astronomical or even infinite search spaces to be explored, so that realistic programs must have more intelligent methods of proving things about themselves. Exploring this may one day teach us what mathematical intuition is.

10.7. Problems about aesthetic experience

Philosophers concerned with the nature of art and aesthetic experience require a theory of perception on which to build. We have seen that from a computational viewpoint, even the simplest forms of perception involve very complex but tightly-interconnected internal processes, which are essentially mental, not physiological, even though we may be largely unaware of them. One way of summarising this is to say that sensory input is like a complex computer program which activates all sorts of different kinds of stored knowledge and abilities, which then interact to generate a process of interpretation which, in turn, may generate other processes, as described in chapter 6 chapter 9 That is we are programmed by whatever impinges on us (see Davies and Isard, 1971). (Of course both people and computers may retain some degree of autonomy in their internal responses to such programming, just as a compiler or operating system does.)

I suggest that aesthetic qualities of experiences are best analysed in terms of the characteristics of these computational processes. Very crudely, a poem, a picture, or tune is more moving, the greater the variety and complexity of the processes it programs. For instance, great music generates processes concerned with auditory experiences, bodily movement, emotional states and intellectual processes including matching structures and resolving ambiguities (Longuet-Higgins, 1976).

Much art and music is shallow because it generates only relatively simple processes or only a restricted range of processes. By contrast, some is shallow because too confusing: the perceptual processes are jammed and fail to activate deeper processes. Occasionally this is because the perceiver needs to be educated. The trade-offs between complexity and power in art are very tricky.

Perhaps one day, in a descendant of the POPEYE program described above, visual experiences will be capable of activating not only stored specifications of general spatial concepts. but also memories of individual past experiences, emotional reactions, and other associations. Designing such systems will give new insights into the process of being *moved* by an experience.

Here are a few further observations about perceptual systems which seem to be relevant to aesthetic issues. Artificial intelligence programs (unlike those in the ’pattern recognition’ paradigm) typically exhibit considerable *creativity* in analysing pictures, understanding sentences, solving problems, etc.

This is because they usually have to work out novel ways of combining their resources for each new task. A picture-analysing program need not have seen a particular configuration previously to be able to interpret it. Often the task of interpreting a picture involves solving some problem (e.g.

Why is there a gap in this line? Which is the best combined interpretation of a group of ambiguous fragments? What are the people in the picture looking at?). We can distinguish pictures according to how complex the problem-solving is, how richly the different sub-processes interact, how many different sorts of knowledge are used, how far it is possible to avoid arbitrary assumptions in arriving at a global interpretation, and so on. These computational distinctions seem to be closely bound up with some aesthetic qualities of a picture, poem or piece of music, often vaguely referred to as unity, harmony, composition, etc. Another issue relevant to aesthetics is the role of different sorts of representation in computer vision systems. See section 10.8. for more on this.

The processes involved in art forms using language (poetry, novels, drama, opera, etc.) are probably more complex and varied than the processes related to painting, sculpture or music. In particular, there is more scope for interaction with huge amounts of knowledge of a whole culture. However, I shall not discuss this topic further.

10.8. Problems about kinds of representational systems

There are several philosophical contexts in which questions arise about the similarities and differences between different forms of symbolism or representation for example in philosophy of mathematics, philosophy of science, philosophy of language and philosophy of art. One of the most important features of artificial intelligence research is the way in which it has generated new sorts of explorations of different forms of representation. In particular two mathematically equivalent methods of representing some collection of information may be quite different in computational power. (This is illustrated in the chapter on learning about numbers, and in the chapter on analogical representations, chapter 7. See also the papers on representations by Hayes, Bobrow and Woods.)

Work on computer vision has included explorations of alternative methods of representation. In particular, although for certain purposes propositional symbolisms are useful, it is often essential that information be stored in structures which to some extent mirror the structure of the image being analysed, or the structure of the scene being depicted. Without this it may be difficult to constrain searches when combining fragments, or checking interpretations for consistency.

Thus programs which do not use analogical representations may take far too long. For instance, a two dimensional array of picture features is often used to reflect neighbourhood relations in the image. Further, in analysing pictures with lots of lines forming a network, it is common to build a network in the computer, representing the topology of the image network. If the image lines depict edges of three-dimensional objects, the very same network can provide a structure from which to start growing a three-dimensional interpretation. Changing the form of representation could seriously affect the time required for certain sorts of processing, even if the same information is available.

Sometimes philosophers discussing the differences between different forms of representation (e.g. Goodman, 1969) suggest that the ease with which we interpret certain sorts of pictures is merely a matter of practice and familiarity. The sort of analysis outlined in chapter 7 shows that this is a shallow explanation, missing the point that there may be important differences in computational power involved. At any rate, all this should undermine philosophical discussions of perception which presuppose that all the knowledge (or beliefs) generated by perceptual experiences can be thought of as propositional, so that questions about the logical validity of inferences arise. For non-propositional representations, non-logical forms of inference, may also be used. Which of them are valid and why, is a topic ready for considerable further investigation. (See also Bundy ’Doing arithmetic with diagrams’ and Brown ’Doing arithmetic without diagrams’.)

10.9. Problems about rationality

More importantly perhaps, instead of merely asking which beliefs, and which rules for inferring beliefs from sense-data, are rational, we can also ask new questions about rationality, such as:

1. Which methods of representation is it rational to use for particular purposes?

2. Are there rational procedures for assessing trade-offs, e.g. trading off increased speed against less economical use of memory space, or increased flexibility against reduced speed or heuristic power against loss of generality?

That is, in the context of trying to design a working person, we see rationality as essentially concerned with processes, strategies, actions and the achievement of goals, rather than with static relations between static objects like sense-data, beliefs or propositions. The Marxist slogan ’The unity of theory and practice’ acquires a new life.

10.10. Problems about ontology, reductionism, and phenomenalism

As remarked previously, much A.I. vision work is anti-reductionist, anti-atomistic. Programs use a variety of concepts from different domains, without any need to reduce them to concepts applicable only to sensory input. Indeed it is arguable that such reductions would generate enormous computational problems. It is much simpler to store and make inferences directly from symbols asserting that one bar occludes another, than to use some translated version mentioning only actual and possible dot-configuration which might depict such a situation. The bar concepts need not even be in principle definable in terms of actual and possible sensory data. All that the system needs is a collection of rules or heuristics for jumping to conclusions about bars on the basis of retinal patterns. The rules need not constitute a definition of ’bar’. This sort of relationship between ’theoretical’ and ’empirical’ concepts is discussed at length in contemporary works on philosophy of science, e.g. Nagel, *The Structure of Science.*

So we see that the artificial intelligence viewpoint provides new weapons for philosophers to use in arguments about phenomenalism and related theories about the nature of perception. More generally: in exploring the problems of designing a robot which can interact with the world, learn things about it, communicate about and reason about it, we are forced to examine the merits of different ontologies. But instead of discussing them in a purely theoretical fashion, as philosophers do, we find that we can put our theories to some kind of practical test. For example, an ontology which leads to a robot that is grossly incompetent at relating to the world is inferior to one which leads to a more successful design. For more discussion on this issue see McCarthy and Hayes, 1969.

10.11. Problems about scepticism

One form of scepticism argues that you cannot ever know that there is an external world containing other people and objects, because a ’malicious demon’ might be fixing all your sense-data so as to deceive you.

Many philosophers have gone to great lengths to try to refute such scepticism in its various forms. I cannot see why, for it is harmless enough: like many other philosophical theories it is devoid of practical consequences.

It is especially pointless struggling to refute a conclusion that is true. To see that it is true, consider how a malicious team of electronic engineers, programmers, and philosophers might conspire to give a robot a collection of hallucinatory experiences. (Even the primitive technology of the 1970s comes reasonably close to this in flight-simulators, designed to give trainee air pilots the illusion that they are flying real aeroplanes.) The robot would have no way of telling that it was tied up in a laboratory, with its limbs removed and its television inputs connected to a computer instead of cameras. All its experiences, including experiences resulting from its own imagined actions, would be quite consistent with its being out romping in the fields chasing butterflies.

Only if it tried some sort of action whose possibility had not been foreseen in the programs controlling its inputs would it get evidence that all was not as it seemed. (Like a flight simulator which cannot simulate your getting out of the plane.)

However, even if you manage to convince yourself that the sceptical arguments are valid, and you have no way of telling for sure that you inhabit the sort of world you think you do, it is not clear that anything of any consequence follows from this. It does not provide any basis for abandoning any of the activities you would otherwise be engaged in. In fact it is only if there is a flaw in the sceptic’s argument, and there is some kind of procedure by which you can establish that you are or are not are not the victim of a gross hallucination, that any practical consequence follows. Namely, it follows that if you care about truth you should embark on the tests.

Since I find it hard to take discussions of scepticism very seriously, I have probably failed to do justice to the problem.

10.12. The problems of universals

How are we able to think of different objects as being of the same kind? Why do we use the same word, for example, ’rectangle’, to describe very many different sorts of objects? What does it mean to say that many objects ’have something in common’? Much philosophical discussion, at least since the time of Plato, has been concerned with these sorts of questions. Answers have taken a wide variety of forms, including:

- the theory that common properties are as much a part of the perceivable world as the objects which have those properties;
- the theory that there is nothing really common to objects which we describe as the same, since we group things together on the basis of arbitrary conventions;
- the theory that such objects have a common relationship to some kind of mental object (e.g. an image or picture with which they are compared);

and no doubt many more.

One of the consequences of trying to give computers the ability to perceive things is that we have to analyse the perception of similarities and differences, and the use of descriptive and classificatory concepts. It seems that the whole thing cannot get started unless there are some kinds of properties and relationships which the sensory system can detect by using measurements or very mechanical (algorithmic) procedures, like matching against templates.

But a real visual system has to go far beyond this in constructing and employing quite elaborate theories as part of the perception process. For example, the program described in the previous chapter has to use the *theory* that one bar partially covers another, to explain a gap in a row of dots in the picture. Less obviously, the ’theory’ that there is a bar in a certain place explains the occurrence of some collinear sets of dots in the sensory image. In view of all the relationships which can be generated by bar-junctions, by occlusion, and by juxtaposition of bars, there is little resemblance or similarity between the different configurations of dots which are interpreted as representing bars at least not enough to distinguish them from others such as configurations which are interpreted as depicting spaces between bars. So using the same label or description for two or more objects may rest on the assumption that they have similar potential for explaining aspects of our experience. So the application of higher-level concepts in describing perceived objects has much in common with the construction of scientific theories to explain experimental results. This sort of point is missed by theorists who try to analyse universals in terms of perceived resemblances or in terms of arbitrary rules or socially determined conventions. (Structuralism, for instance?)

From this standpoint, the particular set of concepts, that is, the set of interpretation procedures and classification rules, used by an animal or person, will probably be the product of a long process of exploration and experiment. The rules which have been most useful in the construction of powerful explanatory theories will have survived. The process of testing such theories involves interacting with the world: moving around, manipulating things, avoiding obstacles, predicting what will be seen from a new viewpoint. This learning need not have been done entirely by individuals: insofar as some mental and behavioural abilities are somehow inherited (for instance, the new-born foal can walk), there is a sense in which *species* can learn though the mechanism of such learning is still a mystery to biologists.

Thus it is to be expected that organisms with partially similar bodies living in a similar environment, will have evolved a not entirely different collection of concepts and theory-building procedures. Such a substratum, common to the whole human species and many animals, might pervade the systems of concepts used in all cultures, contrary to the view that our concepts are essentially *social*, as claimed in the later writings of Wittgenstein and many of his admirers. (Of course, social systems can mould and extend inherited concepts and abilities.)

Further exploration of this sort of idea, in the context of detailed discussion of examples, and the methods by which programs deal with them, will help us transform old philosophical problems, like the problem of universals, into new clearer, deeper problems with which we can make some real progress, and thereby increase our understanding of ourselves.

10.13. Problems about free will and determinism

A common reaction to the suggestion that human beings are like computers running complex programs is to object that that would mean that we are not free, that all our acts and decisions are based not on deliberation and choice but on blind deterministic processes. There is a very tangled set of issues here, but I think that the study of computational models of decision-making processes may actually give us better insights into what it is to be free and responsible. This is because people are increasingly designing programs which, instead of blindly doing what they are told, build up representations of alternative possibilities and study them in some detail before choosing. This is just the first step towards real deliberation and freedom of choice.

In due course, it should be possible to design systems which, instead of always taking decisions on the basis of criteria explicitly programmed in to them (or specified in the task), try to construct their own goals, criteria and principles, for instance by exploring alternatives and finding which are most satisfactory to live with. Thus, having decided between alternative decision-making strategies, the program may use them in taking other decisions.

For all this to work the program must of course have some desires, goals, strategies built into it initially. But that presumably is true of people also. A creature with no wants, aims, preferences, dislikes, decision-making strategies, etc., would have no basis for doing any deliberating or acting. But the initial collection of programs need not survive for long, as the individual interacts with the physical world and other agents over a long period of time, and through a lengthy and unique history extends, modifies, and rejects the initial program. Thus a robot, like a person, could have built into it mechanisms which succeed in altering themselves beyond recognition, partly under the influence of experiences of many sorts. Self-modification could apply not only to goals but also to the mechanisms or rules for generating and for comparing goals, and even, recursively, to the mechanisms for change.

This is a long way from the popular mythology of computers as simple-minded mechanisms which always do exactly what they are programmed to do. A self-modifying program, of the sort described in chapter 6, interacting with many people in many situations, could develop so as to be quite unrecognisable by its initial designer(s). It could acquire not only new facts and new skills, but also new motivations; that is desires, dislikes, principles, and so on. Its actions would be determined by its own motives, not those of its designers.

If this is not having freedom and being responsible for one’s own development and actions, then it is not at all clear what else could be desired under the name of freedom.

As people become increasingly aware of the enormous differences between these new sorts of mechanisms, and the sorts of things which have been called mechanisms in the past (clocks, typewriters, telephone exchanges, and even simple computers with simple programs), they will also become less worried about the mechanistic overtones of computer models of mind. (See also my 1974 paper on determinism.)

10.14. Problems about the analysis of emotions

At various points I have stressed the cognitive basis of emotional states (e.g. in the chapter on conceptual analysis). I have also stressed several times that in an intelligent system there will have to be not just one computational process, but many, all interacting with others. One possible way of analysing emotional states and personality differences, is in terms of different kinds of organisation and control of processing.

For example, my colleague Steve Hardy once remarked that programs which get involved in ’depth-first’ searches, where one of the possible current moves is always chosen, and then one of the moves made possible as a result of that move, and so on, may be described as essentially *optimistic* programs. Similarly, a program which does ’breadth-first’ searches, explicitly keeping all its options open and continually going back to examine other alternatives instead of pushing ahead with a chosen one, could be described as a *pessimistic* program. (The POPEYE program falls somewhere between these extremes.) Of course the program itself is neither optimistic nor pessimistic unless it has been involved in some explicit consideration of the alternative strategies, and has selected one of them. These are simple extreme cases.

Much more complex patterns of control may be involved in a real robot, and by examining different possibilities we can hope to gain new insights into the nature of emotions, moods and the like.

However, it is important to be on guard against superficial computer models. Often by clever programming, people can produce quite convincing displays of something like a mental state, when closer inspection reveals that something very different was going on.
[[This is why the Turing test is of no philosophical significance, since it concentrates only on external behaviour.]]

For example, if hunger, or degree of paranoia, is represented as the value of some numerical variable then that clearly does not do justice to what are actually very much more complex states in people. For example, as anthropologists are fond of pointing out: hunger is not a simple drive to eat. Rather it is a very complex state in which aspects of a culture may be involved. In some communities a hungry person will happily eat caterpillars, locusts, snails, or whatever, whereas members of other communities find such things quite unappetising even when they are very hungry.

More complex desires, emotions, attitudes, etc., involve a large collection of beliefs, hopes, fears, thinking strategies, decision-making strategies, and perhaps conflicts between different sub-processes of the sorts described previously. At the moment, modelling such aspects of the human mind adequately is simply beyond the state of the art. This is why it is sometimes tempting to take short cuts and make superficial comparisons.
[[Note added Sept 2001:
A lot of research in the Cognition and Affect project at the University of Birmingham since I came here in 1991 has been involved in developing the themes of this section. There is a large and growing collection of papers in the project directory http://www.cs.bham.ac.uk/research/cogaff/ including papers challenging shallow behaviourally defined conceptions and models of emotion and contrasting them with architecture-based concepts and theories, e.g.
A.Sloman, Beyond Shallow Models of Emotion, in
Cognitive Processing: International Quarterly of Cognitive Science, 2, 1, pp. 177-198, 2001, available online in postscript and PDF formats.
There are online presentations on these topics in this TALKS directory:
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/
and there is a very flexible software toolkit available free of charge for exploring architectures including architectures in which systems can monitor their performance and modify themselves, described here: http://www.cs.bham.ac.uk/research/projects/poplog/packages/simagent.html **]]**

10.15. Conclusion

This concludes what can only be regarded as a set of notes requiring extensive further discussion. Moreover, the list of headings is incomplete. There are many areas of interaction between philosophy and computing which have not been discussed. Some of them have been mentioned in other chapters. Some, like the theory of meaning (including problems of sense and reference), will have to be discussed on another occasion. Moreover, new points of contact are rapidly emerging. For example, just before finishing this book, I read a review by Meltzer of a PhD. thesis by D. Lenat reporting on a program which explores mathematical concepts looking for ’interesting’ new relationships. The program was able to invent for itself the concept of a prime number and other mathematically important concepts. I have not read the thesis myself, but it is unlikely that the program acquired a very deep understanding of any of the concepts it created. Nevertheless it is still one of the important steps down the long long road to understanding how we work.

If all this succeeds in making most readers want to find out more about A.I., and encourages some people working in A.I. to be more self-conscious about the philosophical presuppositions and implications of their work, then this book will have been worthwhile. I hope a significant subset of readers will be tempted to try *doing* artificial intelligence. This will become easier with the spread of cheaper and more powerful computing facilities, and with the design of improved programming languages. The increasing flow of books and articles on A.I. is also a help. Above all, computers and programming will play an increasing role in educational systems, so that philosophy students of the future will not find the new approach as alien as some of their less well educated tutors do.

At the end of chapter 9, I listed some of the reasons why existing A.I. programs cannot be taken too seriously as models or theories of how people do things. Despite this, the work is essential to the study of how people work (a) because it exposes previously unnoticed problems for instance by showing that even apparently simple abilities depend on very complex computational processes, and (b) because a major obstacle to progress is our lack of adequate theory-building tools, and A.I. research is constantly creating new tools, in the form of new concepts, new symbolisms, new programming techniques, and new aids to exploring and ’debugging’ complex theories. I have begun to illustrate some of the techniques in previous chapters.

Although most of what I have said about A.I. has been concerned with its relationships to philosophical problems, I have also argued that there are strong links with developmental psychology and educational studies. The new insights provided by this sort of work could have a far-reaching effect on a whole range of problems and activities which I have not discussed. For example, in time very many disorders of personality and intellect may be much better understood by thinking of them as involving computational problems (by contrast with regarding them as due to some kind of brain malfunction, to be treated by drugs or surgery, or adopting approaches akin to psychoanalysis without a computational theory to underpin the therapy).

Of course, all this new knowledge might be abused, but it might also lead to great advances in our efforts to help children learn complex concepts, and our attempts to help those whose lives are impoverished by malfunctions ranging from dyslexia to emotional disturbances with a cognitive basis. It is already leading to new advances in teaching techniques, for instance at the Massachussetts Institute of Technology, and the Universities of Edinburgh and Sussex, where new programming languages influenced by languages developed for A.I. are used for teaching computer programming to pupils who previously thought of themselves as bad at mathematics and the use of symbols.[1]

So the title of this book is somewhat misleading. The revolution I have been discussion involves much more than philosophy. The impact of computers and computing on philosophy is merely one facet of a transformation of ways of thinking about complex systems and processes which will increasingly pervade many aspects of our lives and change our image of ourselves. It will thereby change what we are.

Some people regard this as some kind of disaster, and even suggest that the attitude of A.I. researchers and the work they produce can be degrading or dehumanising. For instance, Weizenbaum (1975) comments that when his secretary wished to be left in private while she conversed with a computer, and objected that his plan to record all conversations with his ’Eliza’ program was an intrusion into people’s privacy, he thought that this showed that she was in some sense suffering from a delusion and degrading herself (p. 6). What he apparently did not see is that this is not very different from wanting to be left in private when writing in a book ’Dear Diary .... ’. Suitably programmed computers are much more fun to interact with than a blank page in a book, and the Eliza program is a specially good example.

Moreover the increasing use of computational metaphors for thinking about people is no more degrading than the use of metaphors previously available as a result of advances in science and technology, like the metaphors generated by steam power technology: ’She needs to let off steam’. The pressure built up inside him’, ’He uses music-making as a safety-valve’, ’He was ready to explode’, and so on. The difference is that the new metaphors are richer in explanatory power, as I have tried to show throughout this book. [2]

Endnotes

(1) Of course, in the short run such developments can only have a tiny effect on the mass of the population. Worse, our educational system --- and I include parents, families, churches, prisons, the press, television, and the pronouncements of politicians, in this --- is failing so miserably in so many different ways, that giving everybody a superb grasp of mathematics would still leave much more serious problems: like preparing people adequately for marriage and other personal relationships, making them politically aware and sophisticated, and above all making them thoughtful, considerate, and able to co-operate fruitfully.

(2) After completing this book I read Luria’s fascinating account of *The man with a shattered world*, which shows how brain damage can interfere with some of the processes described in chapters 6, 8 and 9. We now need detailed studies of the links between such clinical phenomena and theoretical speculations about computational mechanisms.

Book contents page
Next: Epilogue

UPDATED:
12 Feb 2009, 26 Sep 2009 (Fixed typos, added note); 19 Sep 2010; 2 Jul 2015 (reformatted)