NOTE ADDED 21 Jul 2015
Since July 2015 this file is out of date.
The completely repackaged book can now be found here in html and pdf versions:
http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp.html
http://www.cs.bham.ac.uk/research/projects/cogaff/crp/crp.pdf

The Computer Revolution In Philosophy (1978)

Aaron Sloman

Book contents page

This chapter is also available in PDF format here.

CHAPTER 4

WHAT IS CONCEPTUAL ANALYSIS?

4.1. Introduction

Elsewhere in this book, I have frequently referred to an activity of philosophers known as conceptual analysis. This has been practised in various forms and for various purposes by a wide range of philosophers and scientists. It has been particularly associated with mid-twentieth-century philosophy in Oxford and Cambridge, for instance the work of L. Wittgenstein, J. Wisdom, J.L. Austin and G. Ryle. As I see it, the main difference between these and earlier philosophers is that the latter were somewhat less self-conscious about the activity. However, on the whole recent analysts agree with previous philosophers that the main function of conceptual analysis is to help clarify or resolve philosophical problems, and occasionally also to provide a basis for criticising some uses of language. For example, in *A Plea for Excuses* Austin claimed that the analysis of the concepts *reason, excuse, justification,* and related concepts would not only help to clarify philosophical problems about freedom but also show some errors in the utterances of judges and in writings on jurisprudence.

I have tried to suggest that, besides these uses, conceptual analysis has another important purpose, namely to find out things about people and the world. However, this requires a far more disciplined and systematic approach to the analysis of concepts than is to be found in the work of most philosophers. (This is partly because their goals are different.)

We have a very rich and subtle collection of concepts for talking about mental states and processes and social interactions, including: *abdicate, abhor, acquiesce, adultery, adore, admire, angry, astonish, attend* and *avid,* to mention a few.

These have evolved over thousands of years, and they are learnt and tested by individuals in the course of putting them to practical use, in interacting with other people, understanding gossip, making sense of behaviour, and even in organising their own thoughts and actions.

All concepts are theory-laden, and the same is true of these concepts. In using them we are unwittingly making use of elaborate theories about language, mind and society. The concepts could not be used so successfully in intricate inter-personal processes if they were not based on substantially true theories. So by analysing the concepts, we may hope to learn a great deal about the human mind and about our own society. This point does not seem to be widely understood: this is why so many people (including many philosophy students) dismiss conceptual analysis as being 'merely concerned with meanings of words'.

Most of the theoretical presuppositions of our ordinary concepts are not concerned with laws or regularities, but with possibilities. For example, the use of a concept like *careful* is based on our knowledge that people can act in certain ways, not on any laws about how they always or usually act. The chapter on the mechanism of mind outlines some results of my own attempts to analyse familiar concepts concerned with actions and related mental processes. These analyses revealed a host of human possibilities, and the mechanism sketched in that chapter is intended to provide the beginnings of an explanation of those possibilities, showing how conceptual analysis can contribute to psychology and artificial intelligence.

Similarly, by analysing concepts related to space and physical motion, e.g. *bigger, longer, inside, push, pull, carry, fetch, throw, impede, collide,* and so on, we may expose some unarticulated theories about our physical environment which govern much of our thought and behaviour. This task is not so urgent because physics and geometry have already made a great deal of progress, often going beyond our common-sense theories. To some extent this has been a result of conceptual analysis: the most striking example being Einstein's analysis of concepts of space and time. However, further conceptual analysis is required for improving our understanding not of the physical world itself, but of how people of various ages and cultures think about the world (consciously and unconsciously). Intelligent machines may need to think of the world as ordinary people do, rather than as quantum physicists do. [Note added: 2001. The recent growth of interest in the study of ontologies in AI and software engineering illustrates this point.]

It has been easier to make substantial progress in the physical sciences partly because the physical world is much simpler than the world of mental and social processes. Moreover, our interactions with the physical world are not as rich as our interactions with people so there is more scope for commonsense to have evolved mistaken theories about matter.

In the rest of this chapter, I shall try to list some of the methods which are useful in analysing concepts. Most of this will be familiar to analytic philosophers, especially those who have studied the work of Austin and Wittgenstein. However,

I have found that the techniques are very hard to teach, and hope that by formulating these procedures, I may help both to clarify how the method works and to provide beginners with a basis for developing the skills involved.

I can only list some techniques for collecting 'reminders' about how our concepts work. The task of organising and explaining the phenomena by means of some kind of generative theory is very difficult. It is similar to the construction of scientific theories. I do not claim to be able to teach people how to be good scientists. (That will have to wait until we have computer programs which behave like good creative scientists, when we shall be in a better position to think about what it is to teach someone to be a scientist!) What follows is merely a sketch, with a few hints. The topic deserves a whole book, and should be susceptible of a better organised presentation than I can manage.

4.2. Strategies in conceptual analysis

When trying to analyse a concept (e.g. *knowledge, truth, emotion, imagination, physical object),* some or all of the following moves may be helpful.

a) Collect descriptions of varied instances of the concept, and also descriptions of non-instances which are similar in some ways to instances. For example, consider the following examples of imagining something: imagining that the Conservatives will win the next election, imagining that you are very rich, imagining that you are falling off a cliff, imagining that there's a donkey in front of you, imagining that time travel will occur one day, imagining that 39875 is the largest possible number. Do these have anything in common? How do they relate to utterances like 'I can't imagine what she sees in him', 'He's a very imaginative dancer', etc. How do they differ? How do they differ from remembering something, learning something, believing something, reading about something, expecting something, planning, and dreaming? Try to formulate rules or definitions which will sort candidates into instances and non-instances, and test your rules or definitions on those previously collected. Try to test them more thoroughly by searching for new difficult cases (which friends and colleagues may be more likely to provide since they will not be committed to your definitions.)

b) Try criticising and extending the definitions given in dictionaries. Dictionary writers are not normally trained in conceptual analysis, and may make mistakes. Moreover, the aim of a dictionary definition is not to explain how a concept works (e.g. what knowledge is presupposed by its use and how it is related to a family of concepts). Rather, the aim is merely to enable someone who already grasps the concept to attach a label to it. So dictionary definitions are usually much too brief and simple to be very useful for analysis of concepts. This comes out most clearly in the attempt to program computers to understand natural language: for such a purpose each word needs to be associated with much more elaborate rules for its use than will normally be found in a dictionary entry. In spite of this, dictionary entries may be good starting points when you are short of ideas.

c) Using a dictionary, and Roget's *Thesaurus,* try to collect lists of related words and phrases: analyses of different items in the same list will probably illuminate each other.

   For example, if analysing the concept *imagine,* look also at *image, imagination, suppose, consider, think, think about, think of, visualise, remember, invent, refer to, have in mind, . . .* Similarly, in analysing the concept *know,* we would need to look at *notice, discover, learn, believe, accept, understand, remember, forget, infer, evidence, reason, test, proof,* and many more. Having found some related but different concepts, try to find examples which illustrate one concept but not the other, and vice versa.

   Try to work out why each example fits one concept but not the other(s). For example, search for examples of knowing X without believing X, or examples of believing X without knowing X. (See Austin's use of examples to analyse the difference between 'by mistake' and 'by accident' in *'A Plea for Excuses'.* My chapter on the mechanism of mind was based on an attempt to extend his work.)

d) Try to collect lists grouped in different ways. For instance, one list given above included mental states and processes related to imagining. Another list would involve *uses* to which imagining may be put, for example, drawing something, solving a problem, trying to recall exactly what happened, entertaining people, anticipating difficulties while making plans, etc. One can then ask how it is possible for the process to be used in these various ways.

This calls for a collection of examples of each kind of use to be thought about carefully, with a view to postulating some underlying mechanism. Another list might include a range of different kinds of things we can imagine (a visual scene, hearing a tune, doing something, a war starting, a mathematical theorem being false, etc.). (One of the things people find hard to learn is the technique of generating examples of things they already know about, including words and phrases. Wittgenstein was a master at this, though he was not very good at analysing the similarities and differences between the examples.)

e) There is a collection of very general categories we use in much of our thought and language, such as: event, act, state of affairs, process, disposition, ability, regularity, cause, explanation, function, object, property, relation, etc. (For example, if you find odd the assertion that apples hang on trees very slowly, this is because you (perhaps unconsciously) recognise that hanging is a *state* whereas 'slowly' can describe only *processes*, like growing.)

Try fitting these categories to the lists of related concepts, to help bring out differences between them. For example, learning something is a process, knowing or believing something a state one is in (perhaps resulting from such a process). Believing something is a state involving a property of oneself, whereas knowing something involves an extra relation to the world (e.g. getting something right). Since knowing something is a state not an event (contrast learning, or discovering, or noticing), those philosophers and psychologists who refer to 'the act of knowing' are either revealing their inability to analyse their own concepts, or else using technical jargon which is bound to cause confusion because of superficial resemblances to concepts from our ordinary language. (I do not wish to deny that ordinary language is itself sometimes muddled.)

Some mental states, for example, believing that there is a tiger in the next room, can explain behaviour, such as running away, but do not involve an ability or any disposition to behave. However, the combination of the belief and another state, such as fear of tigers, may generate a disposition to lock doors, run away, or call for help, depending on circumstances. Some states, for example, knowing how to count, involve an ability which may or may not ever be manifested in behaviour, whereas others, for example, being an enthusiast (e.g. about golf, gardening, or Greek sculpture), involve a tendency or even a regularity in behaviour. 'He smokes' reports a habit which is manifested (much to the annoyance of many non-smokers), whereas 'he would like to smoke' reports an inclination which may be successfully suppressed forever, so that there need not be any behavioural manifestations.

Desiring and wanting are states, whereas deliberating is a process, and deciding an event which terminates such a process and initiates a state of being decided.

Very often noun phrases look as if they denote objects, whereas analysis shows that they do not. Having an image is being in a certain mental state. The state may explain various abilities or actions. Some people think of an image as an object which is somehow involved in the state of having an image much as a nose is involved in the state of having a nose. However, it may be that this is not how the concept works, and that to talk of the image is merely a short-hand and indirect way of talking about a very complex mental state: when we say that a house has a shape we are not saying that besides the house there is some other object, its shape; rather we are alluding to an aspect of the state of the house, namely how all its parts are related to one another.

Similarly if someone has a visual image: this is a matter of being in a state in which one is able to do a variety of things which one can normally do only when there is something one can see. It does not follow that the image is some kind of object like a picture though no doubt, as with all mental states and processes, there is some kind of symbolism used (probably unconsciously) to represent the thing imagined. (For more on this see Pylyshyn, *'What the mind's eye tells the mind's brain").*

f) For each concept being investigated ask whether it refers to a *specific* kind of thing (event, state, disposition, etc.), or whether it covers a whole lot of different kinds of examples, in which case it *is polymorphous* (Ryle, *The Concept of Mind).* For example, the polymorphous concept *motive* covers desires, purposes, attitudes, attempts to achieve something, attempts to prevent or avoid something, and perhaps character traits ('the motive was greed'). If the concept covers many different sorts of cases, this is rarely simply because the word is simply ambiguous. So you can then ask why all these cases are grouped under a common description: do they fulfil a common function? do they have a common explanation? do they have a common relationship to some other things?

   For example, motives have in common the fact that (when combined with beliefs) they can *explain* decisions, intentions, and behaviour. But this shifts the burden to the concept *explain,* or *explanation',* why are there so many different sorts of things we call explanations, and do they have anything significant in common? (An important and still open research question.) *Careful* is another example of a polymorphous concept: different sorts of things are involved in careful driving, careful teaching, careful selection of words in an essay, careful breaking of sad news, careful cleaning of a precious vase, etc. Here it is relatively easy to see what is in common to all these cases, namely reference to goals, possible undesirable occurrences, a collection of risks or dangers, paying attention to the risks, and doing whatever is required to minimise them.

g) If the concept appears to be polymorphous, ask whether there are some 'central' and some 'peripheral' or 'derivative' cases, and try to account for the difference. For example, describing a person as 'moody' or 'unco-operative' seems to be central compared with describing a car that way. Ask what distinguishes central from peripheral or metaphorical cases: is it a difference in the number of preconditions satisfied? If so, why does the concept have those preconditions? What is their point?

h) Ask what the role of the concept is in our culture. Is it merely a convenient descriptive symbol? If so, why should we want to describe those things? Does it have some non-descriptive function? For example, does it express approval? Is its use characteristically abusive, or a means of showing off? Is it part of a system of concepts whose use depends on the existence of some kind of social institution? What is the point of the institution? For example, is it used to apportion blame or responsibility in order to decide questions of redress? What would it be like to live in a culture without that institution? Is there some aspect of the concept which would remain usable without that institution?

   Examples of concepts which seem to depend on more or less complex social institutions are: courage, dignity, disapproval, honour, shame, embarrassment, owning, owing, impertinence and gallantry. Wittgenstein (in his *Philosophical Investigations)* and his followers have argued that very many mental concepts, including 'following a rule', are essentially social. I think that they exaggerate because of their ignorance of possible computational models of mental processes.

i) Ask what sorts of things can be explained by instances of the concept. Does it explain events, processes, states, abilities, non-occurrences, the loss of an ability, success, failure, a single occurrence, a number of occurrences, etc.?

   For example, knowledge explains (or is able to explain) success; fatigue and confusion explain failure; desire explains attempts.

   Does the explanation function as a cause, an enabling condition, a purpose, a justification, an excuse, a mechanism, a law, or what?

j) Ask the following range of questions about instances of the concept under investigation.

1. What sorts of things can bring them about?
2. What sorts of things can prevent them?
3. What sorts of things can facilitate their occurrence?
4. What can cause variations in the instances?
5. What sorts of effects can they have?

Sometimes it is possible to distinguish 'standard' from 'non-standard' causes, effects, etc. For example, there is something irrational about beliefs which are caused by desires ('wishful thinking') but not about actions caused by desires. (Why?)

Sometimes it is useful to distinguish events and processes a person can bring about from those which merely happen. You can decide to stop walking or trying to find something out, but you cannot decide to stop knowing or believing something. You can decide to try to get something, but you cannot decide to want it. Why not? (Answering this question would extend the theory of chapter 6.)

k) If you have managed to collect a number of examples of related concepts, see if you can find a set of relatively 'primitive' concepts and relations, which can be used to generate a lot of the examples, by being combined in different ways. (That is try to find a 'grammar' for the phenomena.) This is a useful first step towards building a good theory of how the concepts work, as opposed to merely describing lots of facts about their relations.

Linguists are increasingly trying to do this though it is not clear how far they appreciate the intimate connection between the study of our language and the study of our world.

For example, the verbs of motion mentioned earlier all seem to involve a subset of the following ideas:

1. Something has a position which changes.
2. Something is an agent (it may or may not also change position, and may or may not change the position of other objects).
3. There is a route for the motion of each object, with a starting and a finishing location.
4. Something may be an instrument, used by an agent, possibly to move an object.
5. Moving things have absolute and relative speeds.
6. If A causes B to move, A may be on the side away from which B is moving or on the side of B to which it is moving.
7. The movement of B may merely be initiated by A (pushing something over the edge of a table) or may be entirely due to A (throwing something, pushing it along).
8. The agent may have a purpose in moving the object.
9. There may be a previous history of movements or locations referred to (e.g. if A retrieves B).
10. There may be more than one stage in the motion (e.g. A fetches B).

11. A may do to B something which tends to produce motion, but the motion may be resisted, e.g. pushing an object which is too heavy, pulling an object with a string which stretches or breaks.

12. The agent may also be the supporter of the object moved, e.g. in carrying it, or may be supported by it, e.g. in riding it.

Different combinations of these (and other) ideas can be used to generate whole families of related concepts, often including concepts for which we do not (yet?) have labels. For example, I do not think English contains a word which refers to a process in which an agent A carries an agent B to some location, and then A picks up some object and is carried, by B, back to the starting point.

Perhaps this is an important part of some social activity in some other culture. Some sort of obstacle race?

The 'primitive' ideas used as the basis for generating such a family of related concepts may themselves be susceptible of further analysis. Moreover, some concepts require mutually recursive definitions: for example, *believe* and *desire* cannot be defined independently of each other.

The sort of analysis suggested here for concepts of motion is now familiar to linguists and people working in artificial intelligence (for example Schank and Abelson, who also explore analogies between such physical processes and mental processes like communicating information. See Bibliography.)

Similarly, in analysing a concept like *know*, or *knowledge*, it will be necessary to distinguish a variety of elements and relations which can enter into scenarios involving knowledge. A person (or other knower) will be involved, as will things in the world about which something is known. There will be a state of mind of the person, in which some aspects of the things and their relationships will be represented, that is, a belief is involved, though not necessarily consciously. There will be something which gives rise to the belief, either at the time the person knows or at some earlier time, for example, a process of perceiving something, doing an experiment or test, or perhaps acquiring the information indirectly from other knowers, or inferring it from some other knowledge.

There will be a relation between the source of the belief and the belief which certifies or justifies the belief (e.g. the evidence is good evidence). There may be sentences, spoken, uttered, or merely thought, which state whatever it is that is known, and in that case the sentences can be decomposed (usually) into fragments with different relations both to items in the world and aspects of the knower's mind. There may or may not be *uses* to which the knowledge is put, including answering questions, interpreting one's experiences, making plans, acting in the world, understanding other people's sentences, formulating new questions, etc. (Again, study of a system of concepts from ordinary language can contribute to psychology, and to the attempt to design artificial minds.)

In two papers on *ought, better* and related concepts (1969 and 1970), I have tried to show how a variety of uses can be generated in a fairly systematic fashion. Similarly, much important work in the development of mathematics, for instance Euclid's, and later Hilbert's, work on the foundations of geometry can be seen as a form of conceptual analysis, though usually of a very reductive sort (that is many concepts and theorems are reduced to a very small number).

l) When analysing a concept it may be helpful to try to list ways in which one can teach a young child or a foreigner learning one's language, how the concept works. What sorts of examples would make good illustrations, and why? What sorts of things would be worth mentioning as *not* being examples, and why are they likely to be confused? What sorts of things need not be mentioned because they can be taken for granted? Why? What would Martians have to be like in order to be capable of learning the concept?

m) Try to list ways in which you can test the truth or falsity of statements involving the concepts in question, including cases which might be difficult. For example, how do we decide whether a person has a certain attitude, such as anti-semitism? Is asking the person an adequate test?

When is it adequate and when not? What patterns of behaviour are adequate tests? Are they *decisive,* or are they merely *indicative?* Why? Are there some situations in which no decisive test is possible, so that doubts cannot be removed? For example, a racialist who has excellent motives for concealing his attitude, and who is an excellent actor. (As we shall see later on, there is no reason to suppose that there should be behavioural tests for all internal computational states and processes, either in a computer or in a person or animal.)

n) Sometimes it is useful to ask whether being in a certain state presupposes having some knowledge, or exercising some intellectual ability. For example emotions like surprise, dismay, embarrassment, shyness and humiliation presuppose a lot of knowledge. You can long for your mother only if you know you have one, know she is not present, and can imagine a possible state of affairs in which the two of you are together. Can a goldfish long for its mother? If not, why not?

The widespread belief within our culture that intellectual and emotional phenomena are quite disparate can be refuted by detailed conceptual analysis.

o) Often some question about the analysis of a concept can be investigated by telling elaborate stories about imaginary situations. So science-fiction writers are good sources of material for this activity. For example, imagine a time when machines are available which will make a complete copy of a human body (including the state of the brain), except that cancer cells are replaced with healthy cells.

Suppose that in such a society it is commonplace for incurable cancer sufferers to agree to have their bodies copied by this machine, while under total anaesthetic, followed by cremation of the cancer-ridden body. The new one is allowed to take its place so people come home from hospital saying I'm glad to be back, and I feel much better now that I've got my new body'. In such a society is our concept 'murder' applicable to their treatment? Is the concept 'same person' applicable to the person who goes into the hospital and the person who comes out? (For more on this see my 'New bodies for sick persons'.)

Another example: people disagree over whether it is essential to the concept 'emotion' that emotions involve felt bodily changes. One way of convincing yourself that such physiological processes are not essential is to imagine a society of Martians who are very much like us with very similar sorts of social institutions and similar ways of seeing, thinking, and acting, but who do not have the bodily reactions which we (or some of us) feel in certain emotional states. So they have hopes, disappointments, pleasant and unpleasant surprises, they feel pity, loneliness, dismay when their plans go wrong, they are anxious when there is a high probability of things going wrong, they are proud of their achievements, envious of others who are more successful, greedy for wealth, and so on. By describing the behaviour and social interactions of such beings in great detail, and imagining what it would be like to communicate with them, you should be able to convince yourself that you would find it perfectly natural to use our emotion concepts in talking about their mental states.

You would say 'He's terribly embarrassed about the attention he's getting', even though he feels no hot flush in the cheeks or any other physiological change characteristic of embarrassment in humans.

Of course, this sort of investigation does not produce knock-down arguments, because people can differ in how their concepts work. For example, mathematicians use a concept of *ellipse* which includes circles, whereas for non-mathematicians a necessary condition for something being an ellipse is that it has major and minor axes of differing lengths. Similarly, there may be some people for whom the accompanying physiological changes are necessary conditions for the applicability of concepts like *envy, embarrassment, loneliness,* etc. However, what one can demonstrate to such people is that by insisting on these necessary conditions they are making it impossible for themselves to describe situations which might one day arise, without inventing a whole lot of new terminology which may prove very hard to teach. Whereas I would claim that my use of the non-physiological concept of emotion in no way interferes with my communication with other people, and allows me the power to read science fiction without any feeling of linguistic distortion.

p) Try to test your theories by expressing them in some kind of computer program or at least in a sketch for a design of a working program. For example, try to design a program which can communicate with people using the concepts. If you have analysed the concepts wrongly then this will show up in some failures of communication between the computer and people (just as the misunderstandings of children and other learners show up). Or test your analysis by designing a program whose behaviour is intended to *instantiate* the concept, then see whether the actual behaviour is aptly described using the concepts in question. You will usually find that you have failed to capture some of the richness of the concept. For example, for a while some people hoped that programs written in the language PLANNER would capture the essence of the concept of a goal, or purpose. But the behaviour of the programs clearly quashed this hope. (E.g. see Winograd, 1972.)

   Of course, sometimes a little thought makes this elaborate kind of test unnecessary. Nevertheless, the methods of A.I. provide a useful extension to previous techniques of conceptual analysis, by exposing unnoticed gaps in a theory and by permitting thorough and rapid testing of very complex analyses.

This account of conceptual analysis is by no means complete. For more detailed examples, refer to the writings of philosophers mentioned and also A.R. White's *Attention,* and his contribution to *Owl of Minerva,* (ed. Bontempo and Odell), and Margaret Boden's *Purposive Explanation in Psychology.* Philosophers usually do not pay enough attention to problems of describing mental processes. Neither do they normally attempt the kind of system-building involved in designing a 'grammar' for a collection of concepts in the manner hinted at above. For instance, is there some sort of grammar for concepts related to attention? In other words, is there a relatively small subset of concepts in terms of which all the others can be defined? I believe the answer is 'Yes' but to establish this will require designing a fairly detailed model of a person, capable of generating a large number of processes involving perception, deliberation, reasoning, planning, problem-solving, and execution of plans and intentions. Some small steps in this direction are taken in Chapter 6, which proposes some minimal architectural requirements for a human-like system.

Despite my disparaging remarks about philosophers, there have been some profoundly important systematic analyses, mostly produced by philosophers of logic and mathematics, such as Frege, Russell, Tarski and Prior. For example, Frege's analyses of concepts like *all, some, nobody,* and related quantifiers, led to a revolution in logic and has profoundly influenced the development of computer programming languages used in artificial intelligence (via the work of Alonzo Church). Austin's *How to do Things with Words* is another example of a philosopher's attempt at detailed and systematic analysis, which has made a great impact on linguistics and more recently on AI.

If only Wittgenstein, in his later writings and teaching, had not made such a virtue of his inability to construct systematic theories integrating the results of his analyses, a whole generation of philosophers might have been far more disciplined and productive.

Of course, there are dangers in insisting on everything being formalised and systematic. Much shallow theorising is a result of trying to fit very complex and messy structures into a neat and simple formal system. A well known example of the distorting effect of formalisation is the claim that the logical connectives of propositional calculus adequately represent the words 'and', 'not', 'or', 'if, etc. of ordinary language. However, even if this claim is false, it remains true that the formalisation provided a basis for deeper exploration than was previously possible. For example, by describing exactly how the use of the ordinary words deviates from the truth-functional symbols, we obtain useful descriptions of how they work. (See Gazdar and Pullum 1977.) The same can be said of some other systematic but inaccurate analyses.

The two extremes to be avoided are demanding formalisation of everything at all costs, and rejecting formalisation because some of our concepts are too complex and unsystematic in their behaviour for us to be able to represent them in elegant formal systems. One of the great advantages of using programming languages for formulating analyses of concepts (as Winograd did see his 1973), is that programming languages are well suited to include many tests for special cases and exceptions to general rules. It is much harder to use formal grammars, or axiomatic systems, for this purpose.

4.3. The importance of conceptual analysis

The activity of attempting to analyse families of related concepts can be enjoyable and interesting in its own right. Discussion of similarities and differences between *fetch, retrieve, carry,* and related concepts is the sort of thing even a child can find good fun though getting the analysis right is not child's play. But besides giving intellectual pleasure, the activity may have a useful function. For example, it is well known that many perennial philosophical problems arise out of confused reflections on things we all know, and that at least some of these problems can be solved or dissolved with the aid of conceptual analysis. I think it can also be shown that a great many debates on ethical and political issues, such as debates about the justifiability of abortion, about equality of educational opportunity, and about what sorts of decision-making procedures are democratic, are often more confused than necessary either because the participants are using concepts in a muddled fashion or because they are to some extent at cross purposes because of subtle differences in the ways their concepts work. In either case progress can be made if people learn how to analyse their own and other people's concepts.

Conceptual analysis can play a role in science and mathematics too. I have already mentioned Einstein's work involving analyses of concepts like *simultaneous,* and other spatial and temporal relations. Another example is the struggle by mathematicians of previous centuries to clarify the concepts *infinite* and *infinitesimal,* leading to the discovery of the concept of a limit, and to formal set theory.

Every science will have at its frontiers concepts which are to some extent in need of analysis and possibly improvement. Not all the problems of science are to be solved simply by collecting new facts, or by using existing terminology to build new theories. In the mature sciences, the concepts most in need of analysis will usually be highly technical, remote from the concepts of ordinary language.

However, in the social sciences and psychology, and increasingly in artificial intelligence, concepts from ordinary language play a central role in the construction of new theories and in the description of phenomena to be explained. Thus it is important for practitioners of these disciplines to be sensitive to the need for analysis, and to be skilful at doing it.

The dangers of failing to analyse concepts properly can be illustrated by a few rather extreme examples. Someone who had not seen how the concept *bachelor* worked might think it interesting to do a survey to find our what proportion of the bachelors in some social group were unmarried. He would probably get no support from research councils. However, less obvious mistakes of the same sort could pass unnoticed, like attempts to test the hypothesis that *other things being equal* people tend to believe things which are asserted by those they respect, or the hypothesis that *other things being equal* people tend to try to achieve goals they think they can achieve, or the hypothesis that being embarrassed involves believing that other people are paying attention to you. Of course, such research goals would usually be disguised in obfuscating jargon, but that does not reduce the need for conceptual analysis. I once read a research proposal which looked very impressive until the English equivalent to the jargon emerged. The aim was to find out whether people tend, on the whole, to co-operate more successfully if they get on well together. (For some similar criticisms of Social Science, see Andreski, *Social Science as Sorcery.)*

An example of an important piece of biological theorising whose concepts cry out for detailed analysis can be found in Dawkin's *The Selfish Gene.*

Besides the role of conceptual analysis in preventing muddled thinking and silly research, there is another important role in relation to science, namely making explicit some of what we already know, clearly a useful preliminary to attempts to add to what we know. I believe this is especially useful in fields like developmental psychology and anthropology, concerned with the study of ways of thinking and learning. Previously I listed some concepts concerned with spatial movement and indicated how one might begin to analyse some of the more complex ideas in terms of combinations of relatively primitive ones. Very young children somehow acquire both the relatively 'primitive' concepts and also a variety of complex combinations of these. It is not thought to be beyond them to grasp the difference between 'fetch' and 'send for' expressions which occur in familiar nursery rhymes. By studying these concepts we can define some of the tasks of psychology. An adequate theory of learning must account for a child's ability to master these ideas. Even very young children are capable of grasping quite abstract rules, including rules which they cannot formulate in words. For example, a three-year-old reacted to his older brother's use of 'nope' for 'no', by starting to say not only 'nope' but also 'yesp', 'okayp' and 'thankyoup'. Try formulating the rule he had invented! (Do developmental psychologists, or brain scientists, have any convincing explanation of the ability to learn these things?)

By improving our understanding of what it is that our children have to learn we may perhaps come to understand better not only how they learn, but also what sorts of things can go wrong with the learning process, and perhaps even what can be done about it. How many teachers in schools, colleges and universities have sufficient skill in conceptual analysis to be able to discern subtle differences between the concepts they are trying to teach and the concepts so far grasped by their pupils?

Other social sciences can also benefit from conceptual analysis. By doing this sort of analysis for concepts used in several different cultures, anthropologists and sociologists could enhance their studies of what is common and what varies among different modes of thinking and reasoning.

I have already alluded several times to the role of conceptual analysis in the work reported in this book. Several chapters are based in part on attempts at analysing familiar concepts. But most of the work is still sketchy and makes use of concepts which themselves require further study.

The chapter on the aims of science, for example, makes liberal use of a very complex concept which still requires further analysis, namely the concept of what is *possible.* Several other concepts used in that chapter are equally in need of further investigation.

The chapter on analogical representations attempts to analyse a familiar distinction between different sorts of symbolisms, or representations, showing that the verbal/pictorial distinction is usually misdescribed and that there are actually several different distinctions where at first there seems to be only one.

The chapter on learning about numbers begins to analyse some of our simplest number concepts, drawing attention to complexities in what a child has to learn which are not normally noticed.

The chapter on computer vision, and the ensuing discussion includes some small steps towards clarifying a collection of familiar concepts like *conscious, interest, experience.*

Nearly all of this work is incomplete, and will remain incomplete for many years. But, as I have suggested in this chapter and will try to substantiate later, the methodology of artificial intelligence will be a major spur to progress.

[[Note Added November 2001
Since this chapter was first published, the problem of 'knowledge elicitation' in designing expert systems has received much attention. It is not widely appreciated that the techniques of conceptual analysis as described here (and practised by many philosophers) are often crucial to such knowledge elicitation. There is also considerable overlap between these ideas and the Naive Physics project proposed by Pat Hayes: See P.J. Hayes, The second naive physics manifesto, in *Formal Theories of the Commonsense World* Eds. J.R. Hobbs & R.C. Moore, Norwood, NJ, Ablex, 1985, pp. 1-36

Note added February 2007
Additional discussion of the nature of conceptual analysis, its relationship with what Gilbert Ryle called 'logical geography', and a possibly deeper notion of 'logical topography' can be found here http://www.cs.bham.ac.uk/research/projects/cogaff/misc/logical-geography.html **]]**

Endnotes
(1) Margaret Boden, Frank O'Gorman, Gerald Gazdar and Alan White commented usefully on an earlier draft.

Book contents page
Next: Chapter five
Last updated: 4 Jun 2007; Reformatted 1 Jul 2015