5 THE COGNITIVE APPROACH

Thus far, we have talked of cognitive structures and cognitive processes. Section 3 offered some examples of historical proposals as to what kinds of things cognitive structures and processes are. Contemporary cognitive psychology equates representations with cognitive structures, and computations over these with cognitive processes.

5.1 Representation

We have emphasized the scientific nature of cognitive psychology. However, Fodor (1974) argued that psychology might be a special science – special because its subject matter, the mind, stands in a complex relation to the material, physical world – and therefore takes a different form from the natural or social sciences. Spelling out the relationship between the mind and the physical world, even between the mind and the body, is extremely difficult. Two competing intuitions have guided people's thinking about the issue. One is that the mind transcends the physical body (and the brain) – that when we say we are in love, for example, we mean more than that we are in a particular bodily or brain state. Though you may share this intuition, it is difficult indeed to say what a psychological state is if it is *not* physical. It is also difficult to reconcile this intuition with the methods of natural science – how is it possible to study something scientifically if it is not physical in nature? The competing intuition is that all aspects of humanity, including our minds, ought to be explicable as parts of the natural world, and so explicable by the natural sciences. Humans are, after all, products of natural, evolutionary pressures, shaped by the world in which we have evolved. How could we come to possess a mind that could not be explained as part of the natural, physical world?

The tension between these two intuitions is real and difficult to resolve (as you will see from Chapters 18 and 20). Here we can do no more than hint at the difficulties. One feature of the mind may go some way towards showing why the intuitions are so difficult to reconcile. It is the feature of representation.

Some things in the world have the property of being 'about' something else. Books, for example, tend to be about other things. A book on the history of Rome is about precisely that – the real events that go to make up the history of Rome. The observation is so mundane that you may never have given it a second thought. Yet this property of **aboutness** is quite extraordinary, and certainly difficult to explain within the natural sciences. A book, for example, could be described physically in terms of the arrangements of its molecules, the kinds of atoms that it comprises, its chemical compounds. We could describe its mass and volume, and measure it for electrical and magnetic properties. Yet these descriptions produce no hint as to a book's subject matter. Only when the patterns of ink are considered, not as patterns of ink, but as *words*, does it become clear what a book is about.

Few, if any, things in the natural world have this property of aboutness. It makes no sense to ask what a stone is about or what a river is about. While it makes sense to ask what a book or a newspaper is about, it makes no sense to ask what its components, the ink and paper, are about. It *does* make sense to ask what mental or cognitive processes are about – we often say to one another, 'What are you thinking about?' One way of expressing the aboutness of mental processes is to say that they involve **representations** – our thoughts *represent* possible states of affairs, our perceptions *represent* our immediate environment (generally, though not always, accurately).

The representational quality of mental processes was described by the philosopher of psychology Franz Brentano (1838–1917). Brentano believed that mental states comprise mental *acts* and mental *contents*. So, for example, my believing that Rosie, my pet cat, is lazy is a mental state – I am in the state of believing that Rosie is lazy. For Brentano, the state has a dual character: it comprises an act, corresponding to the act of believing, and a content, namely the content that Rosie is lazy. Brentano thought that mental states can differ, even if they involve the same mental act. So, for example, my believing that Rosie is lazy, and my believing that all cats are lazy, would represent two different mental states. The same act is common to both, but the beliefs are differentiated by their content: one is *about* Rosie; the other is *about* all cats.

The consequence for Brentano was that psychology needs to consider not only the internal features of the mind or brain, but also what these features are about or represent in the world. Perhaps now you can see why it is not straightforward to decide what kind of science cognitive psychology is. Whereas physics and chemistry study the material world of atoms and molecules (which do not have this representational quality), cognitive psychology studies mental states whose representational nature cannot be ignored. Consequently, cognitive psychology studies something intrinsically relational – something that spans what is in the mind and what it relates to in the world. Indeed, the issue of representation tends to distinguish the social sciences (such as sociology) from the natural sciences (like physics). Cognitive psychology, focusing on both what is represented (the world) and what does the representing (the mind), does not fall neatly into either category.

5.2 Computation

In Section 3 we considered some of the technological and theoretical antecedents to cognitive psychology. What emerged from the advances concerning theories of information and computation was the view that computers process information and provide a means for modelling and understanding the mind. As David Marr put it, 'If . . . vision is really an information processing task, then I should be able to make my computer do it . . .' (Marr, 1982, p.4).

Marr's statement hints at a deep relation between the computer and the mind. If computers process information, and information processing is what characterizes minds, then perhaps, at some deep level, the mind is computational. This claim provides a further key assumption of the cognitive approach: cognitive psychologists tend to view the mind as computational, as well as representational.

Von Eckardt (1993) suggests that there are two assumptions involved in construing the mind as computational. First is a linking assumption – the assumption that the mind is a computational device of some kind and that its capacities are computational capacities. The assumption serves to link minds (things that we wish to understand better) with computers (things that are already well understood). Second is the system assumption – this fleshes out what is meant by a computational device. Generally, the assumption tends to be that computers are systems that represent information, input, store, manipulate, and output representations, and which operate according to rules. The two assumptions work together to provide a framework for understanding the (relatively) unknown mind in terms of the known computer.

Just as with the representational assumption, the assumption that minds are computational raises many questions. One of the more pressing for cognitive psychology has been the precise form that computational models should take. This is in fact a major debate within contemporary cognitive psychology, and the issue will be referred to in one way or another in many chapters in this book (especially in Chapters 19 and 20). Broadly speaking, there have been two main proposals as to the computational models we should use to understand the mind: symbolic models and connectionist models.

5.2.1 Symbol systems

One way of understanding the idea that the mind is both representational and computational has been to suggest that the mind is a symbol system. In this view the representational qualities of the mind are expressed via the claim that the mind is symbolic and contains symbols. So, for example, my mental state that Rosie is lazy might be described as involving symbols for Rosie and laziness. The symbols together represent what the belief is about. To say that the mind is computational is to say none other than that the mind embodies (computational) mechanisms for manipulating these symbolic representations. My believing that Rosie is lazy would then involve my appropriately manipulating the symbol for Rosie and the symbol for laziness.

Newell and Simon (1976) were the first to propose that the mind is a symbol system. In their view, symbolic representations and their manipulation are the very building blocks of intelligent thought and action. Newell and Simon proposed many different properties of symbol systems but we need consider only a few. Symbol systems should comprise a basic set of symbols that can be combined to form larger symbol structures (just as the symbols for 'Rosie' and 'lazy' could be combined to form the symbolic expression 'Rosie is lazy'). Symbol systems should contain processes that operate on symbol structures to produce other symbol structures. Finally, symbol structures should represent, or be about, objects.

Newell and Simon's proposal that the mind is a symbol system amounts to the claim that the cognitive processes that underlie language, perception, memory, thinking, categorization, and problem solving will ultimately turn out to involve processes of manipulating and transforming symbolic representations. The proposal is, of course, an empirical one, and in principle the evidence could turn out either way. One way of addressing the issue is to develop models of symbol systems and compare these with empirical data (e.g. from human participants in an experiment). As you will see throughout this book, the strategy of producing computer models and comparing their performance with human data is a common one (see especially Chapter 19 for such comparisons for symbolic models). However, it is worth noting that disagreement with empirical evidence does not necessarily imply that the cognitive processes in question are not symbolic. It may well be that a different symbolic model would agree with the data much better. So, although the claim that the mind is a symbol system is empirical, it will require a considerable amount of empirical evidence to show either that the mind is symbolic or that it is not.

reverse engineering Attempting to understand how a pre-existing device works on the basis of how it behaves.

Marr's levels of explanation and cognitive psychology

How, therefore, can such information processing devices be best understood? To answer this, we must now turn to a framework for thinking that was provided by Marr (1982). According to him it is possible to discuss three levels of description of any kind of information processing device, namely:

1. The level of the computational theory.
2. The level of the representation and the algorithm.
3. The level of the hardware implementation.

The level of the computational theory

At the **computational theory level**, concern is with what the device does and why it does it. It is at this level that the 'logic of the strategy' (Marr, 1982, p. 28) is spelt out. Consider the following example of an electronic calculator and the question of how it carries out arithmetic. Analysis at the level of the computational theory would address the fact that the calculator carries out various arithmetic operations (the 'what it does') and the fact that it uses a particular method for carrying these out (the 'why it does what it does'). For instance, an early Hewlett Packard calculator (the HP35) used a method based upon something known as Reverse Polish. So expressions such as:

$(1 + 2) \times 3$

were entered as

$1\ 2\ 3 \times +$

which accords with something known as postfix notation (or Reverse Polish – and because of this, and perhaps the price, it is no wonder this model quickly died out). The computational theory would be therefore be concerned with issues like why Reverse Polish was used and what the principles of Reverse Polish are.

The level of the representation and the algorithm

At the **representation and the algorithm level** much more detailed questions are asked about the nature of the calculator's operating system and the manner in which numbers and arithmetic processes are embodied in the device. In this respect, we are interested in how information is stored (i.e., represented) within the device and also how various arithmetic operations are instantiated. We will discuss in much more detail the notion of representation as we proceed through this book (see Chapter 7, for instance). For now, though, we will stick with a simple definition and assert that for any information processing device, information from the outside world is represented internally within the device. So when the number 2 is entered into the calculator this is represented via some form of electronic code. This form of **internal representation** stands for the number 2. By analogy, where the mind is concerned such internal (mental) states stand for (i.e., represent) actual states in the real world.

What then happens to such representations is a matter for the **algorithm**, or more particularly, the set of operations that are carried out on these representations. In computer science the term 'algorithm' is mainly used interchangeably with the phrase 'computer program', but we may take a more general reading and define it as a procedure that, when correctly applied, ensures the correct outcome. If you entered a '+' sign into the calculator and it is working properly then it should invoke its addition algorithm. The addition algorithm comprises the sequence of operations that determine that two numbers are added together. So understanding the nature of the calculator depends on trying to specify the nature of internal representations and the associated internal algorithm. In terms of understanding human cognition, and by analogy, we need to consider both mental representations and mental processes. In this respect our functional account should not only provide a flow chart that maps out the relations between component processes, but also some description of the sorts of internal representations that are also implicated. A much more thorough exploration of these ideas is contained in the next chapter, but in summary, at the level of the representation and the algorithm we are committed to being precise about (i) how states of the world are represented by the device, and (ii) what the contingent internal processes are.

The level of the hardware

Finally there is the **hardware implementation level** and, as we have already noted, flow charts such as that shown in Figure 1.6 are of little use. Concerns at this level are with how the designated representations and processes are implemented physically. What physical components do we need to build the device? As has been discussed, the one purpose of a functional description is to avoid any commitment to physical instantiation, but of course to attain a full understanding of any physical device details at all three levels of explanation will need to be addressed.

Pinpoint question 1.9

According to Marr (1982), what are the three levels of description for any device?

Levels of explanation and information processing systems

Let us examine the notion of levels of explanation in more detail and apply this in an attempt to understand the operation of a programmed computer: the paradigm case of an information processing system. One way of conceptualising how a computer operates is in terms of the following levels:

1. The intended program.
2. The actual computer program.
3. Translation.
4. Machine instructions.
5. Transistors.

The level of the intended program – Level 1 – is where discussion of what it is that the program is designed to do takes place. Using Marr's framework, this is couched at the computational level and may be fleshed out by stating what we want the program to do – we need to specify the goals and objectives behind the program. For instance, we want the program to produce a bank statement that itemises all in-goings and out-goings on an account in a month. At this point there is no need to make any reference to any specific kind of computer or hardware of any kind. At the next level (i.e., at Level 2) the necessary step is to commit ideas to 'paper' by writing a computer program in a particular language, e.g., BASIC, PASCAL, FORTRAN, etc. In completing these two stages both the level of the computational theory and the level of the representation and algorithm have been captured. Importantly, both of these levels have been addressed without any concern whatsoever about the nature of the computer that the program will run on.

For computer languages to be at all useful, however, there needs to be a means of converting (i.e., translating) an actual program into a form that can be run on a particular computer. Stated thus, concerns about hardware are paramount. In some programming languages the translation stage may involve two steps: (i) converting the computer program into an intermediate code known as assembly language, and then (ii) converting assembly language into machine instructions. It will be something of a relief to learn that we need not concern ourselves with the detailed nature of assembly language.

The critical point is that the stage of translation takes a representation of the program and converts it into something known as binary code – a series of 0s and 1s where each 0 and 1 is known as a bit. For instance, the command PRINT might be translated into 0101000010101010. The 0s and 1s are vital because they correspond to the respective OFF/ON states of the critical electronic components of your computer – here referred to as transistors. Think of these as being akin to simple switches in an electric circuit: if the switch is ON, electricity flows this way round the circuit; if the switch is OFF, the circuit is closed. So there is a fundamental level at which properties of the program correspond exactly with physical states of the machine. The machine code corresponds to the physical state of the computer's 'switches' and these cause the computer to behave in particular ways. Any change in state of the switches results in the computer doing something else.

This example provides a concrete illustration of how Marr's levels of analysis may be useful in attempting to explain the operation of a computer. The central idea is that the computer can be described at a number of different levels, and that there is no sense in which a description at any particular level is more correct or valid than a description at any other level. The different descriptions serve different purposes. If the questions are about what the program is supposed to do, then these concern the level of the computational theory. If the questions concern the design of the program, then these will be answered most appropriately at the level of the representation and the algorithm. If the questions concern whether the program uses an 8-bit or 16-bit representation of numbers (don't worry, your computer science friends will be only too delighted to tell you what these are), then the answers will be at the level of the machine code and therefore at the level of hardware implementation.

Research focus 1.3

What's a computer? Half a century playing the imitation game

You are sitting in front of a computer, idly instant messaging. But how do you know the person on the other end is really another human being? What if the human responses were occasionally replaced by computer-generated responses? Do you think you would be able to tell? This proposal is a variation of the famous imitation game, now commonly referred to as the Turing Test first proposed by Alan Turing in 1950, and as summarised by French (2000). In it, Turing was attempting to define the sufficient conditions under which a computer could be attributed with thinking and intelligence – two of the most human of all cognitive faculties.

While the relevance of the Turing Test is doubted by some, French (2000) stated that the imitation game will continue to raise important questions as the bridge between human and machine becomes closer, such as: 'To what extent do machines have to act like humans before it becomes immoral to damage or destroy them?' (pp. 116–17). Anyone who cradles their MP3 player in a rather loving way might appreciate the sentiment here.

The Turing Test has been criticised as being a rather behaviourist approach to the definition of intelligence, in that all the machine needs do is succeed in the game, and this does not really constitute an adequate conceptualisation of what it is to think (see Gunderson 1964, cited by French, 2000; and, Searle, 1980), while others have argued that to 'play the game' is not a single ability but one that requires numerous cognitive faculties.

Despite the Turing Test being proposed over 50 years ago, it still provokes intense debate. This is perhaps most represented by the Loebner Prize, which offers a $100,000 reward for the first person who can design a computer that passes the Turing Test. French (2000) stated, however, that few computer programs have even come close and that the prize might actually impede serious discussion of the philosophical issues arising from the test itself. So the next time you're instant messaging, it might be worth thinking about who might be on the other end.

Source: French, R. M. (2000). The Turing test: The first 50 years. *Trends in Cognitive Science, 4,* 115–122.

Levels of explanation and reductionism

Having considered how it is that computers operate, it seems odd to try to argue that one level of explanation is better or more valid than any other since different questions demand answers at different levels. Nevertheless, in questioning the utility of having more than one level of explanation we are beginning to consider what is known as *reductionism* – the doctrine that we can reduce all accounts of behaviour to a very basic physical level (perhaps even to the atomic or even subatomic level – see Penrose, 1989). If we accept some version of central state identity theory, then we are accepting that mental states and processes are nothing other than neural states and processes. According to reductionists, understanding the mind reduces to understanding the basic electro-chemical states and processes that characterise the behaviour of neurons (brain cells). If we understand these physical states and processes, we understand the mind.

A more virulent version of reductionism is known as *eliminative materialism*. Churchland (1984) states, essentially, that once we have a full understanding of the behaviour of neurons in terms of basic electrochemical principles, we can then eliminate any mention of any other level of description from our science. Mental states and processes reduce to neural states and processes, and if we understand neurons then we understand the mind. Churchland (1984, p. 44) is particularly taken with eliminative materialism because, on his reading of history, this sort of theorising has been successfully applied in ridding science of unnecessary concepts such as 'phlogiston' (i.e., a spirit-like substance that had been assumed to be given off when materials burn or metal rusts). The argument is that, given that eliminative materialism worked for physics, why can't it work for cognitive psychology? Maybe we can get rid of the mind after all?

This line of argument is very threatening to those of us who study cognitive psychology because it seems to suggest that as soon as we have a complete understanding of the brain, there will be nothing left for cognitive psychologists to do. More importantly, it seems to lead to the conclusion that mind science should be completely replaced by brain science! However, this conclusion can be resisted and a reasonable rebuttal is that even if you did have a complete account of the nature and operation of neurons, this is couched at a completely different level to the one that we as cognitive psychologists are concerned with. Indeed, such an understanding would have little utility if we want to make predictions about the everyday life of human beings. For instance, we may conjecture that the primary reason that Harry went to the florist's and bought his girlfriend a bunch of flowers was because he thought she was still angry at him after she caught him gazing longingly at her best friend. This is not to argue that such a physical (neuronal) account is not possible, in principle, but just that it would be of little use in trying to understand Harry's concerns about trying to figure out his girlfriend. These sorts of considerations lead to the conclusion that both reductionism and eliminativism are misguided (for a much more technical exploration of these ideas see the introduction of Fodor, 1975). On the contrary, what we have argued is that mental states and processes can only be understood if we have some understanding of their functional roles. As cognitive psychologists we are interested in uncovering the functional nature of the mental components that constitute the mind – the flow charts of the mind – and in this regard the properties of neurons will be of little help.

In this respect Marr's (1982) framework for thinking is so important. In using his kind of analysis the characteristics of any information processing device can be laid bare. It is also assumed that in using his analysis the relation between the mind and the brain becomes tractable. For cognitive psychologists the levels of the computational theory and the level of the representation and the algorithm come to the fore. This is not to dispute the fact that considerable progress continues to be made in human neuroscience. The main point is that we, as cognitive psychologists, focus on the two levels – which are essentially the cognitive levels – that do not concern anything that might be going on at the level of neurons. In the same way that the very same program can run on two different computers, we assume that the same kinds of mental representations and processes are embodied within different brains. In this way cognitive psychologists can (and do) operate without any regard for the brain.

Pinpoint question 1.10

How does reductionism threaten cognitive psychology?

computational theory level Marr's (1982) first level of description is concerned with what the device does and why it does it.

representation and the algorithm level Marr's (1982) second level of description concerned with how a device represents information and how such representations are operated upon.

internal representation Some property of the inner workings of a device that stands for some property of the outside world. For instance, in a computer 01 represents '1'.

algorithm A well-specified procedure (such as a computer program) that ensures the correct outcome when correctly applied.

hardware implementation level Marr's (1982) third level of description concerned with the physical nature of an information processing device.

Concluding comments

There is no doubt that we have covered quite a bit of territory in moving from behaviourist beginnings to contemporary theorising in cognitive psychology. Along the way we have established that there is a basic method to the madness of trying to understand the abstract entity that is the human mind. If we can make testable predictions about mental events in terms of some form of measurable behaviour, then it is quite legitimate for us to study mental events. Such a claim, however, must be approached with some caution and we have been mindful of the problems in attempting to accept a complex cognitive account when competing simple behavioural accounts exist. When should we appeal to Occam's Razor and when should we not?

We have also noted with some concern a desire to reduce our cognitive theories to the level of neural processes on the assumption that once we understand the brain we will understand the mind. Again we have expressed caution here. It seems that where complex information processing systems are involved, it is probably better to adopt a more circumspect approach and accept that such devices can be understood at various levels. Neither level is more valid or correct than any other level.

Responding to the Chinese room argument

The Chinese room argument has been much discussed by philosophers and cognitive scientists. My aim here is not to come out for or against the argument. It is simply to introduce you to some of the main moves that have been made (or might be made) in the debate – and so to to give you the tools to make your own assessment of its power and plausibility.

Many people have pointed out that there seems to be a crucial equivocation in the argument. The physical symbol system hypothesis is a hypothesis about how cognitive systems work. It says, in effect, that any cognitive system capable of intelligent behavior will be a physical symbol system – and hence that it will operate by manipulating physical symbol structures. The crucial step in the Chinese room argument, however, is not a claim about the system as a whole. It is a claim about part of the system – namely, the person inside the room who is reading and applying the instruction manual. The force of the claim that the Chinese room as a whole does not understand Chinese rests almost entirely on the fact that this person does not understand Chinese. According to what Searle and others have called the *systems reply* to the argument, the argument is simply based on a mistake about where the intelligence is supposed to be located. Supporters of the systems reply hold that the Chinese room as a whole understands Chinese and is displaying intelligent behavior, even though the person inside the room does not understand Chinese.

Here is one way of developing the systems reply in a little more depth. It is true, someone might say, that the person in the room does not understand Chinese.

Nonetheless, that person is still displaying intelligent behavior. It is no easy matter to apply the sort of instruction manual that Searle is envisaging. After all, using an English dictionary to look words up is not entirely straightforward, and what Searle is envisaging is more complex by many orders of difficulty. The person inside the room needs to be able to discriminate between different Chinese symbols – which is no easy matter, as anyone who has tried to learn Chinese well knows. They will also need to be able to find their way around the instruction manual (which at the very least requires knowing how the symbols are ordered) and then use it to output the correct symbols. The person inside the room is certainly displaying and exercising a number of sophisticated skills. Each of these sophisticated skills in turn involves exercising some slightly less sophisticated skills. Discriminating the Chinese characters involves exercising certain basic perceptual skills, for example.

A supporter of the systems reply could argue that we can analyze the ability to understand Chinese in terms of these more basic skills and abilities. This would be a very standard explanatory move for a cognitive scientist to make. As we have seen on several occasions, cognitive scientists often break complex abilities down into simpler abilities in order to show how the complex ability emerges from the simpler ones, provided that they are suitably organized and related. This is the source of the "boxological" diagrams and analyses that we have looked at, included the Broadbent model of attention (in section 1.4) and the Petersen model of lexical processing (in section 3.4). A cognitive scientist adopting this strategy could argue that the system as a whole has the ability to understand Chinese because it is made up of parts, and these parts individually possess the abilities that together add up to the ability to understand Chinese.

Searle himself is not very impressed by the systems reply. He has a clever objection. Instead of imagining yourself in the Chinese room, imagine the Chinese room inside you! If you memorize the instruction manual then, Searle says, you have effectively internalized the Chinese room. Of course, it's hard to imagine that anyone could have a good enough memory to do this, but there are no reasons to think that it is in principle impossible. But, Searle argues, internalizing the Chinese room in this way is not enough to turn you from someone who does not understand Chinese into someone who does. After all, what you've memorized is not Chinese, but just a complex set of rules for mapping some symbols you don't understand onto other symbols you don't understand.

Exercise 6.9 How convincing do you find this response to the systems reply?

Another common way of responding to the Chinese room argument is what is known as the robot reply. We can think about this as another way of developing the basic idea that we need to analyze in more detail what understanding Chinese actually consists in (in order to move beyond vague intuitions about understanding or its absence). Some writers have suggested that the Chinese room, as Searle describes it, is far too thinly described. The problem is not with what goes on inside the room, but rather with what goes into the room and comes out of it.

A supporter of the robot reply would agree with Searle that the Chinese room does not understand Chinese – but for very different reasons. The problem with the Chinese room has nothing to do with some sort of impassable gap between syntax and semantics. The problem, rather, is that it is embodied agents who understand Chinese, not disembodied cognitive systems into which pieces of paper enter and other pieces of paper come out. Understanding Chinese is a complex ability that manifests itself in how an agent interacts with other people and with items in the world.

The ability to understand Chinese involves, at a minimum, being able to carry out instructions given in Chinese, to coordinate with other Chinese-speakers, to read Chinese characters, and to carry on a conversation. In order to build a machine that could do all this we would need to embed the Chinese room in a robot, providing it with some analog of sensory organs, vocal apparatus, and limbs. If the Chinese room had all this and could behave in the way that a Chinese-speaker behaves then, a supporter of the robot reply would say, there is no reason to deny that the system understands Chinese and is behaving intelligently.

Again, Searle is unconvinced. For him the gulf between syntax and semantics is too deep to be overcome by equipping the Chinese room with ways of obtaining information from the environment and ways of acting in the world. An embodied Chinese room might indeed stop when it "sees" the Chinese character for "stop." But this would simply be something it has learnt to do. It no more understands what the character means than a laboratory pigeon trained not to peck at a piece of card with the same character on it. Interacting with the environment is not the same as understanding it. Even if the Chinese room does and says all the right things, this does not show that it understands Chinese. The basic problem still remains, as far as Searle is concerned: simply manipulating symbols does not make them meaningful and unless the symbols are meaningful to the Chinese room there is no relation between what it does and what a "real" Chinese-speaker might do.

Exercise 6.10 Explain the robot reply and assess Searle's response to it.

Clearly there are some very deep issues here. Searle's arguments go right to the heart, not just of the physical symbol system hypothesis, but also of the very question of how it is possible for an embodied agent to interact meaningfully with the world. Searle sometimes writes as if the problems he raises are specific to the enterprise of trying to build intelligent symbol manipulators. But it may be that some of his arguments against the physical symbol system apply far more widely. It may be, for example, that exactly the same questions that Searle raises for the robot reply can be asked of ordinary human beings interacting with the world. What exactly is it that explains the meaningfulness of our thoughts, speech, and actions? Some cognitive scientists have given this problem a name. They call it the *symbol-grounding problem.* It is the subject of the next section.