2. Which language do we think with?

I have accepted Fodor's view that propositional attitudes (beliefs, desires, and so on) are best understood as relations to sentences. The question then is: which sentences? In this chapter, after first considering the evidence from psychology (which is equivocal), I shall present an intuitive, introspection-based, argument for the view that human conscious thinking involves sentences of natural language. This is an initial presentation of the view that I shall wish to defend in this book. I shall then begin to consider Fodor's arguments for claiming that the sentences in question belong to an innate, universal, language of thought (Mentalese). Some of these arguments are easily responded to, but some will require extensive discussion in later chapters.

2.1 The evidence from scientific psychology

What does cognitive psychology tell us about the relationship between language and thought? I shall consider the developmental evidence first, and then briefly present and discuss the evidence from neuropsychology, particularly aphasia. I shall argue, in summary, that the scientific evidence is inconclusive, remaining open to a variety of interpretations.

The evidence from normal development

It is an oft-remarked fact that the linguistic and cognitive abilities of young children will normally develop together. If a child's language is advanced, then so will be its abilities across a range of tasks; and if a child's language is delayed, then so will be its cognitive abilities. To cite just one item from a wealth of empirical evidence: Janet Astington (1996) reports finding a high correlation between language-ability and children's capacity to pass false-belief tasks, whose solution requires them to attribute, and reason from, the false belief of another person. (A standard false-belief task has the form: Maxi places some chocolate in a cupboard and leaves the room; in his absence we move the chocolate to another
location; question: where will Maxi look for the chocolate when he returns? Most three-year-old and autistic children fail this task, saying that Maxi will go to where the chocolate actually is. Almost all four-year-olds pass, saying that Maxi will look in the cupboard; see Wellman, 1990, and Perner, 1991.) Does this and similar data show that language is constitutively involved in children's thinking? By no means.

As we pointed out in section 1.3, the most that such data can reveal is that language is a necessary condition for certain kinds of thinking, not that it is constitutively involved in those kinds. And this is easily explicable from the standpoint of someone who endorses the communicative – as opposed to the cognitive – conception of language. For language, in human beings, is a necessary condition of normal enculturation. Without language, there are many things that the child cannot learn; and with delayed language, there are many things that the child will only learn later. It is only to be expected, then, that cognitive and linguistic development should proceed in parallel. It does not follow that language is itself actually used in the child's cognition.

But what of the fact that a child's language-development will often lag behind its conceptual and cognitive development? Does this serve to show, on the contrary, that thought is independent of language? For example, when observed (and video-taped) in context, a two-year-old who can only use two-word sentences like, 'Mummy shoes', may sometimes mean, 'Those are Mummy's shoes', (a possessive statement), and may sometimes mean, 'Mummy, put my shoes on', (a request to an actor to perform an action on an object – see Gregory, 1987, p. 427). Then since the child's capacities for thought go beyond what it is capable of expressing in language, does this show that thought is independent of language? In fact not, for two distinct reasons.

First, the argument would move too swiftly from the child's deficiency in language production to the more general conclusion that it is deficient in language-ability. But this inference is not warranted. For it may be that the language faculty divides into two distinct sub-systems, one concerned with production and one concerned with comprehension. (Indeed, the evidence from aphasia suggests that this is the case. Some aphasics are very poor at producing intelligible discourse themselves, but have quite good comprehension abilities.) And if language is implicated in thought, then it may be that it is (or can be) the comprehension system which is so implicated. So if the child is capable of understanding the difference between the two utterances, 'Those are Mummy's shoes', and 'Mummy, put my shoes on' (said by a sibling, say), then it may be that it thinks to itself, 'Those are Mummy's shoes', in circumstances in which it only has the ability to say aloud, 'Mummy shoes'.
Second (and even if the above point is waived), it does not follow from the data that language is not directly implicated in the child’s thought. From the fact that the child makes use of one-and-the-same (two-word) sentence, now to mean, ‘Those are Mummy’s shoes’, and now to mean, ‘Mummy, put my shoes on’, it does not follow that it can think those thoughts independently of the use of that (or an equivalent) sentence. It is possible that the child’s ability to make use of signs outstrips its ability to mark those different uses by different signs — but that its thoughts are partly constituted by the signs so used.

The evidence from abnormal development

If the evidence from normal development is equivocal, then what of the evidence from cases where language development is wholly absent, such as the examples of ‘wolf-children’, or pre-signing deaf children? When these were discussed in section 1.3, it appeared that they, too, presented no real evidence either way. For we may grant that the cognitive development of these children is severely impaired; but this datum is consistent with either the cognitive or communicative conceptions of language. From the standpoint of the cognitive conception, the reason why their cognition is so impaired is because more complex forms of human thinking actually involve the use of language; so where language is absent, so must these forms of thinking be. But from the standpoint of the communicative conception of language, the reason why these children are cognitively impaired is that they lack many normal human concepts and beliefs, which can only be acquired through enculturation and linguistic communication.

But what of the fact that the isolated child Genie, when she was first discovered at age 13, had a mental age of between 1 and 2 — despite then going on to display intelligence in the normal range when she later acquired language? (see Curtiss, 1977). Is this immense cognitive deficit wholly explicable in terms of the absence of communication, and so in terms of Genie’s lack of enculturated beliefs? Perhaps not. But then there are many other features of Genie’s case which might explain the differential, consistent with the communicative conception of language. For it was not just communication that she was deprived of.

Genie was also deprived of all normal emotional contact with others, and of opportunities to explore her environment (which was in any case very impoverished). Almost her only contact with other people, between the ages of 1 and 13, were occasional visits from her father and brother, neither of whom spoke to her, the former of whom beat her regularly (especially if she made any sound), and both of whom tormented her and...
barked at her like a dog. She was also kept immobile for much of her life, almost all of which was spent in a dim, largely unfurnished room, either strapped to a commode, or zipped into a specially adapted sleeping bag, which effectively functioned as a strait-jacket. It would hardly be surprising if, in the circumstances, Genie’s cognitive development was very severely delayed, way beyond what one might predict from her lack of opportunities to receive information from others. And, indeed, the pre-signing deaf man described by Schaller (1991) displayed deficits which were not nearly so severe.

Schaller describes the case of a profoundly deaf man, Ildefonso, who reached the age of 27 without acquiring any conventional language. He had been brought up in a community of migrant labourers, without any schooling, and without any contact with natural sign languages. He was balanced, alert, and normally affectionate, and capable of using and understanding simple gestures and uses of mime. He was capable of performing everyday tasks, and of earning a wage as a labourer. When introduced to ASL (American Sign Language) he learned it extremely fast, sucking in each new word hungrily. Schaller herself describes Ildefonso’s predicament as one in which the absence of language had deprived him of a great deal of information, thus committing herself to the communicative conception of language. But actually, hardly any formal testing of Ildefonso’s cognitive abilities was undertaken, either before and after language acquisition.

Moreover, it is doubtful whether Ildefonso was genuinely languageless. Evidence from elsewhere suggests that he would have developed for himself a gesture system having all of the properties of a simple language. (Indeed, this interpretation seems confirmed by Schaller’s description of him engaging in animated ‘conversations’ with other non-signing deaf adults, using what she describes as a rich repertoire of gesture and mime.) Susan Goldin-Meadow and colleagues have studied deaf children born of hearing parents over a number of years, whose parents then took a decision that the children should not be taught any form of Sign (see Goldin-Meadow and Mylander, 1990, Butcher et al., 1991, and Goldin-Meadow et al., 1994). They found that without exception the children spontaneously developed a gesture language, initially taking over the gestures of their parents, but then systematising and regularising them into a genuine language, with all the properties of morphology and syntax which one would expect of simple three-year-old language. So it seems that it may actually be impossible to find any genuinely languageless humans, unless, like Genie, they are deprived of human contact altogether!

Does the fact that congenitally deaf children spontaneously develop a gesture language show that they are already capable (in advance of
Which language do we think with? Possessing any language of entertaining complex thoughts? For certainly the invention of an entire language (albeit a simple one) can be no simple matter! But this is almost certainly not the right way to understand what takes place. For in fact the languages of these children pass through exactly the stages of development and spontaneous restructuring as do those of children acquiring ASL; and surely no one wants to say that in the normal case children have to invent their own language! (But see section 2.8 below.) Rather, it seems that the human language-forming faculty not only contains a powerful set of innate constraints (as Chomsky and others have argued; see Chomsky, 1988, Curtiss, 1989, and Pinker, 1994), but that it requires only minimal linguistic exposure to be triggered into growth and activity. This same interpretation is also strongly suggested by the way in which children brought up of pidgin-speaking parents will spontaneously elaborate them into Creoles in the space of a single generation (see Holm, 1988). So there is no real reason to believe that young children are capable of complex thought prior to the development of natural language – for languages are, in fact, grown rather than learned or invented.

Of course, the fact that Ildefonso probably already had some language does not show that he had nothing more to learn. And interestingly, Schaller reports that when she began to teach him to sign, he had particular problems in understanding and learning to use signs for temporal concepts. It took months of intensive training for him to begin to use tenses, for example. This certainly suggests that the process of learning to sign was not merely a matter of acquiring labels for a set of pre-existing concepts. Yet it is difficult to see, either, how temporal concepts might presuppose a set of beliefs which can only be acquired through the use of language after extensive enculturation. Acquiring the capacity to think about the past and the future is not a matter of acquiring a whole body of beliefs about the past and future, within which temporal concepts are embedded. On the contrary, it is, surely, inseparable from acquiring competence with some system of temporal markers, or symbols.

So does it then follow that temporal concepts, and the thoughts containing them, constitutively involve the use of public signs? Not necessarily, in fact. For it may be that there are concepts which can only be acquired through the use of language, but where the concepts so acquired are thereafter independent of language. That is, it may be that Ildefonso’s initial temporal thoughts (considered as tokens, not types) were inseparable from the natural language signs which constitutively expressed them. But it may be that, once competence was acquired, he was thereafter capable of thinking thoughts of the same type without the use of those public signs. For example, competence in the use of public-language
temporal signs, once attained, may be mapped across on to an equivalent system of Mentalese symbols, enabling temporal thoughts to be entertained independently of public language thereafter. So once again there is nothing in the data to force us to embrace the cognitive conception of language (at least in its strongest form). It would appear that the evidence from abnormal, as from normal, development is equivocal.

The crucial evidence which would seem to be needed here is this. What happens to a congenitally deaf child who is developmentally ready to acquire full-blown language (at age 5 or 6, say), but who has had no previous exposure to conventional sign-language, if they are then immersed in a signing community? One would predict that they would learn to sign extremely fast. If they were immediately to display evidence of dramatic changes in the creativity and flexibility of their thinking, across a range of cognitive domains, then this would count strongly in favour of the cognitive conception of language. For it would surely be implausible that at this stage the children could have had the opportunity to acquire many new beliefs through enculturation. Indeed, much of the early use of language involves comment upon items that are already perceptually present to both adult and child, and so the child cannot be acquiring any new (non-meta-linguistic) beliefs. The best explanation would be that these items of vocabulary, and the linguistic system of which they form part, would make it possible for the child to engage in a new kind of thinking - namely (as I shall argue), conscious thinking.

The only evidence bearing on the above proposal that I know of, arose out of the experiment conducted by Luria and Yudovich (1956). They took a pair of five-year-old twins who had, through neglect and lack of motivation resulting from the self-sufficiency of the ‘twin situation’, acquired barely any language. They separated the twins and placed them in environments where they would have the motivation to learn to communicate, and, in addition, subjected just one of them to a course of language-training. They reported marked improvements in cognitive ability with the beginnings of acquisition of language (after just three months, in fact). When first discovered, the twins appeared incapable of even rudimentary forms of play. They would move objects about and line them up, say, but would make no attempts at construction or planning, let alone at symbolic play. They also appeared incapable of classifying heterogeneous collections of objects into kinds. But as their language improved, so their play became more ordered, thoughtful, and creative; and similarly their classification abilities became more normal.

The results of this experiment are extremely suggestive for the thesis that some kinds of thinking actually involve natural language, but by no means conclusive. One initial source of worry is that the experiment was
conducted within a theoretical framework which assumes the involvement of language in thought (following Vygotsky, 1934). Luria and Yudovich's interest was not so much in testing this framework, as in elaborating some of its details. This meant that alternative explanations of their data were neither proposed nor tested. It is therefore hard, now (especially at this distance in time), to have full confidence in the reliability of their results. And so far as I am aware the experiment has never been repeated.

It might also be objected that the twins' acquisition of language was not the only factor which may have increased the sophistication of their play. For their separation meant that they were no longer self-sufficient, and had to seek to interact with, and join in the play of, the other children in the kindergarten. Their improved play-performance may then be attributable to observational learning from the play of normal children, rather than to increased language ability. But this objection can be met, in fact, by appealing to the differential performance of the two twins after their initial separation. After three months, the language-abilities of the twin who had undergone special training were markedly better than that of his sibling, as were both the creativity of his play and his classification abilities. And this differential cannot be put down to increased opportunities for observational learning, of course.

There remains, however, the possible rejoinder which parallels the one we recently made in the case of Ildefonso – namely, that language may be necessary for the acquisition of certain concepts and modes of thinking, without necessarily thereafter being implicated in those thinking. This proposal is hard to refute from the developmental evidence alone. For almost everyone accepts that language-acquisition is instrumental in the child's acquisition of new concepts (the notable exception being Fodor, 1981, who maintains that all concepts are innate). But it is quite another matter to maintain that the concepts so acquired are thereafter tied to the deployment of language.

So, much of the twins' dramatic cognitive development could be put down to their acquiring new conceptual structures resulting from their acquisition of language, without us necessarily having to allow that natural language was actually implicated in the use thereafter made of those conceptual structures – that is to say, in their thinking. It could be, in particular, that the twins' growing competence in the use of natural-language logical operators (particularly the conditional) was responsible for their increased creativity and improved planning abilities. In which case natural language may have been constitutively involved in their early token conditional thoughts. But it is another matter to claim that it remained constitutive of those thoughts as a type. It may be that once competence was attained it was transferred to some Mentalese equivalent
of the conditional, enabling the twins to entertain conditional thoughts thereafter in the absence of natural language.

The evidence from aphasia

I turn, now, to consider the evidence from aphasia. This is a condition in which people who were previously normal have lost their capacity for language, resulting either from a head injury or from a stroke. The first point to be made is that there are many different forms of aphasia, each of which may admit of various degrees of severity. There are aphasias which only affect speech output, either removing it altogether, or reducing it to strings of meaningless jargon. There are aphasias which affect comprehension, aphasias which affect word-recall, aphasias which affect grammatical ability, and so on. The most interesting population for our purposes consists of so-called ‘global aphasics’, in whom both production and comprehension are severely affected.

Unfortunately, many investigations of intelligence in aphasia have been population-studies making little attempt to distinguish different types of aphasia (see Zangwill, 1964, and Wyke, 1988, for reviews). Moreover, severity of aphasia has often been measured by degree of impairment of language output, rather than by impairment of comprehension. Such aphasics are, for obvious reasons, more convenient test subjects, since they can understand and respond to instructions. And many researchers may still have been influenced by the account of inner speech due to the behaviourist John Watson, who analysed it in terms of action inhibition (see ch. 10 of his 1924, reproduced in Lycan, 1990). For on this account, any disruption in the capacity to generate speech would at the same time disrupt inner speech, hence also disrupting thought, on the hypothesis that thinking is conducted in language. But there seems no good theoretical reason for assuming that inner speech cannot be conducted in auditory imagination; and, indeed, there is a good deal of empirical evidence that it actually is (see Gathercole and Baddeley, 1993). Even Steven Pinker (who should know better) is to be found arguing from the spared cognitive abilities of a patient with production-aphasia, to the conclusion that thought is independent of language (see his 1994, ch. 2). Needless to say, no such conclusion is warranted.

It must therefore be cases of severe global aphasia which are of most direct relevance to the question of the involvement of language in thought. Unfortunately, however, it is by no means easy to devise non-linguistic tests of cognitive ability. (Witness the difficulties experienced by primatologists in devising a non-linguistic version of the false-belief tasks; see Gomez, 1996.) However, Raven’s Matrices, some of which are rather
like jigsaw puzzles, have been widely used. Andrew Kertesz (1988) summarises the data by remarking that aphasic performance in these and other non-verbal tasks is generally impaired, with the degree of impairment correlating significantly with the degree of comprehension deficit. But he also notes the existence of severe global aphasics in whom a variety of non-verbal abilities are quite well preserved.

Even if there is some cognitive ability which is never found to be left intact in cases of global aphasia, however, it would not necessarily follow that the normal operation of that ability actually involves the use of language. For it may be that the basis of the ability in question is sited in the same region of the brain as the language system, without either of them being functionally implicated in the operation of the other. It might then be that any brain-injury which damages the one will necessarily damage the other, without the one actually involving the other.

For example, many studies report some form of disorder of conceptual thinking associated with aphasia – that is, disorders which involve failures to classify objects appropriately, or to notice semantic connections between pictures, say (for reviews see Lebrun and Hoops, 1974, and Wyke, 1988). Does this support the view that deployment of concepts is tied to the use of lexical items of natural language? Not necessarily, for the reason just given. It would certainly make sense, after all, that conceptual structures (of Mentalese, by hypothesis) should be stored and processed in an area of the brain very close, at least, to the language centres, given that so much of their normal deployment is through language, by engagement in language production or language comprehension. If the essential function of natural language is the communication of thought, as the communicative conception of language maintains, then it is perhaps only to be expected that language should be represented in the brain very close to the thought-components (concepts, lexical items of Mentalese) which it has the function of communicating.

Conversely, even if there are clear cases of dissociation between global aphasia and some cognitive ability (such as spatial reasoning abilities), it does not follow that the ability in question is not language-involving. For it may well be that the language faculty consists of three distinct sub-systems, and not just two. (This is Chomsky’s view; see his 1995. He believes that in addition to the production and comprehension systems, there is a distinct knowledge-base which underpins both.) It may well therefore be that, besides a production system and a comprehension system, there is also what might be called a linguistic inference system. This would have the function of manipulating sentences, inferring other sentences from them, and so on. And it may be that this – thinking – system can be left intact when both production and comprehension systems are
2.2 The evidence of introspection

destroyed. (Note that Chomsky appears to commit himself to the possibility of such an inference system, when he postulates a level of representation in the language faculty which operates upon logical form, independently of phonological encoding; see Chomsky, 1995, and May, 1985.)

The issue of the extent of the involvement of language in thought has been much debated amongst aphasiologists, with expert opinion on the question being divided. (The terminology employed is striking: those who maintain that language is involved in thinking are known as ‘sensualists’, while those who maintain the independence of thought from language are ‘spiritualists’. It pleases me that the thesis being defended in this book makes me a sensualist who is opposed to spiritualism!) As I understand it there is, even now, no emerging consensus on the issue, with many aphasiologists turning to investigate other, more tractable, matters.

While evidence from scientific psychology may eventually prove decisive in resolving the question of the involvement of natural language in thought, and would certainly be necessary to plot out in any detail precisely which kinds of thinking constitutively involve such language (as we shall see in Chapter 8), it is too early, as yet, for any firm conclusions to be drawn from the available scientific data. I shall therefore turn to consider whether we might fare better by contemplating a very different set of data, namely that delivered by ordinary introspection. It may be that, in looking to the evidence from science, we have been conducting our search too far from home, and that the case for the involvement of language in thought is actually right under our noses, available to consciousness.

2.2 The evidence of introspection – images and imaged sentences

If introspection is to be used to support the involvement of language in thought, then it is important that the question we direct to ourselves should be properly framed. Otherwise it can easily seem that introspection lends support to the sort of imagist theory of thinking which we discredited in section 1.8. For if the question I ask myself is, ‘What do I mean by the word “car”?’ (for example), and if I then say the word to myself while I introspect, what I find, almost inevitably, is that I have formed a visual image of a car. But the mistake here consists in asking for the meaning of a word in isolation – as Gottlob Frege pointed out long ago (see his 1884). Rather, the meaning of a word consists in its contribution to the meanings of whole sentences containing it, and a word only properly has its meaning within the context of a sentence.
So what one needs to do is, firstly, to introspect while (or shortly after) using some sentence of natural language in the course of one’s daily life; and secondly, while (or shortly after) one has been entertaining privately some completed thought, or sequence of such thoughts. In the first sort of case what one discovers (as I have already indicated in the points made in section 1.8 above, following Wittgenstein) is that there is often no separable mental process accompanying the utterance of the sentence itself; or, at least, not one that is available to consciousness. In the second sort of case what one discovers, I believe, is that our private thoughts consist chiefly of deployments of natural language sentences in imagination – inner thinking is mostly done in inner speech.

In this second sort of case there is a limited body of systematically gathered introspective evidence available. Russ Hurlburt has devised a method for sampling people’s inner experience (see Hurlburt, 1990 and 1993). Subjects wear headphones during the course of the day, through which they hear, at various intervals, a randomly generated series of bleeps. When they hear a bleep, they are instructed to immediately ‘freeze’, then record and elaborate, what was passing through their consciousness at the time. Although frequency varies widely, all normal subjects report experiencing inner speech on some occasions (with the minimum being 7% of occasions sampled, and the maximum being 80%; most subjects reported inner speech on more than half of the occasions sampled). Most subjects also sometimes report the occurrence of visual images, and the occurrence of emotional feelings (between 0% and 50% of occasions sampled in each case). (Moreover, most people also report entertaining some wordless, or purely propositional, thoughts. This is a finding I shall return to in section 2.4 below, and then again at some length in section 8.3.)

Introspection informs us, in fact, that many of our thoughts are expressed entirely in natural language – for example, my thoughts as I write this, and, no doubt, yours as you read it. When I sit and draft a letter to someone in my head, for example, what figures in my consciousness is a sequence of English sentences in auditory (and perhaps kinaesthetic) imagination, rather as if I were dictating that letter aloud. Or, if this is considered too easy for me (since the task is itself a linguistic one), take as an example any case where I sit and try to think through some abstract problem – comparing competing explanations of autism, say. In such cases I find that my thoughts will consist almost entirely of inner dialogue. Consider another point, too – learners of a second language will characteristically report that there is a watershed in the learning process at which they begin to be able to think in (and even dream in!) the language under study. Then such people, at least, must be
under the impression that they do their thinking in one natural language or another.

Many others of our thoughts consist of combinations of natural-language sentences and mental images. For example, while driving home I might be considering how to re-arrange the furniture in my office, and I might think to myself, ‘I shall put the desk there’, where there relates to a visual image of a particular location by the window. Even in those rarer cases where the only introspective vehicle of thought is a mental image, it seems that there is, almost always, an implicit sentential embedding to endow the image with a determinate content. For example, when faced with a particular practical problem, such as how to reach an object on top of a tall wardrobe, the solution may come to me simply in the form of a visual image, such as myself standing on top of a stool. In such a case the image may get its significance from my disposition to entertain in thought a sentence such as, ‘If I place the stool like that, then I shall be able to reach.’

Introspection informs us of a good deal more than all this, indeed. Some people report that their thoughts occur to them in the form of images of different heard voices, as if a number of different people were speaking to them, or as if they were hearing themselves speak (see Hurlburt, 1990 and 1993). For example, without prior teaching or prompting I asked my son Isaac (then aged 4½) how he did his thinking. He replied, ‘I think in English.’ I asked him how he knew this, and he replied, ‘I can hear myself think.’ As for myself, I can report that most of my thoughts occur in the form of imaged conversations, in which I imagine myself saying things, often, though by no means always, to a determinate audience.

In fact this is the grain of truth, it seems to me, in the imagist theory of thinking. It is images of natural language sentences which are the primary vehicles of our conscious thoughts, and we think and reason primarily by manipulating such images. So, while it may be true that we could not think consciously at all unless we had imagination, studying the nature of mental imagery will be a good deal less useful in investigating the nature of thought than will studying the nature of natural language. For it is not the image, as such, which carries the content of the thought, but rather what is imaged – namely, a natural-language sentence.

We often do our thinking aloud, or on paper. (Indeed, this fact causes particular problems for the communicative conception of language, with its special emphasis on the role of natural language in communication, as we shall see in Chapter 3.) It is common for young children to accompany their games and activities with spoken monologues, and many adults, too, will chat to themselves when alone, or to their babies and pets. Even more
frequently, those of us whose work involves the written word - whether academics, or bureaucrats, or business executives - will do much of our thinking on paper, or at the keyboard. In many such cases one does not first entertain a private thought and then write it down; rather, the thinking is the writing.

Even many overtly communicative uses of language can equally be characterised as a kind of public thinking. For example, consider cases where one reasons with someone, trying to convince them of the truth of some proposition, or of something that they should do. Here, again, the activities of thinking and speaking seem inseparable. All this is provided, of course, that introspection is correct in telling us that we think in the public medium, rather than that we translate our prior thought into that medium.

According to introspection, then, private and public thought are alike in that they both involve sentences of natural language. Just as a good deal of private thinking consists in the imaging of spoken or heard sentences, and in the manipulation of such images; so, too, many overt uses of language would seem to constitute a sort of public thinking. And both sorts of thinking operate in an essentially similar manner - namely, in the one case in the imaging of, and in the other case in the utterance or writing of, sequences of natural language sentences.

These similarities may be further confirmed by our impression of what happens when we make the transition from private thought to public utterance, or vice versa. For example, I may be thinking, privately, about where to go on holiday next year, entertaining such thoughts as, ‘It would be nice to go somewhere hot.’ Then, realising that my wife has entered the room, I may say aloud, ‘We could go to Avignon this time.’ In such cases there is no introspective impression of hiatus, or of transition between radically distinct symbolic systems. It is quite unlike even those cases where one makes a spoken transition between different natural languages - as when, while in France, I negotiate a purchase for my son, and have to switch backwards and forwards between English and French. Indeed, it seems much more like those instances which often occur when I lecture, where I utter one sentence aloud and then write the next on the blackboard. In both cases the sentences follow one another smoothly, as part of the same unbroken chain of thought.

2.3 The scope and strength of the introspective thesis

The arguments sketched above seem to show that at least some human thought does in fact involve natural language - though I shall shortly begin to consider some possible replies from the lovers of Mentalese. But
the scope of the thesis is, so far, quite severely restricted. In the first place, of course, the thesis defended above extends only to thoughts that are conscious, since it was based on the deliverances of introspection. If there are thoughts that are non-conscious, as I shall suggest in Chapter 5 that there are, then nothing has as yet been said to suggest that they, too, will involve natural language. Perhaps less obviously, but more importantly, the introspective thesis so far defended only extends to occurrent, or episodic, acts of thinking, not to thoughts that are standing states, as are most beliefs, desires, and intentions.

There is a general distinction to be drawn amongst mental states between those that are episodic events — occurring at particular times with particular, if not exactly measurable, durations — and those that are not. In the former category will fall bodily sensations, emotions, perceptions, imaginings, and after-images, as well as acts of thinking, wondering whether, supposing, deciding, remembering, and so on. When I feel a stab of pain in my ankle while running, or see a rabbit jump from my path into the bushes, or wonder whether it is time to turn for home, these are all mental events that occur at particular times in my life.

In the category of standing states, in contrast, will fall beliefs, desires, hopes, intentions, and memories, as well as character traits such as generosity and credulity. My belief that Sheffield is in England, and my memory of my fifth birthday-party, are states that I will retain throughout most of my life. The general term ‘thought’, in philosophical use, includes events and states from both categories, covering all those that take propositional objects, being canonically described by means of a that-clause. So perceptions, propositional thinkings, imaginings, and acts of remembering, as well as long-term beliefs, desires, and intentions (but not sensations or character traits) are all of them thoughts, in this sense.

Now it is only episodic thoughts that can be characterised as events in consciousness, or that are immediately available to introspection. For one can, for example, easily forget what one believes, or desires, or intends. Since we cannot directly introspect our beliefs, desires, or intentions, it has not yet been shown that they, too, involve natural language. So the most that has been established by the introspective argument given above, is that our episodic acts of thinking involve natural language, not that all thoughts do.

Even this is too strong, indeed. For in philosopher’s usage, not only would acts of judging-that, wondering whether, and such like, be described as ‘(occurrent) thoughts’, so too would perceptual states of seeing-that, hearing-that, and so on, as also would states of imagining-that. Yet it is hardly very plausible to claim that we see in natural language, or entertain visual or auditory images in natural language! When I notice
a car approaching while I am crossing a road, and quicken my pace accordingly, it does not seem plausible to claim that my perception of the car is expressed in English, or constitutively involves any English sentence. Of course I *may* think consciously to myself, 'I had better get out of the way of that car'; but then I may not – I may just act appropriately, on the basis of my conscious perception. Equally, as we noted in section 1.8, when I entertain a visual image of the street-layout of the city of Sheffield, it is plain that I need not be entertaining any English description.

So the most that I should wish to claim on the basis of introspection, is that it is episodic, properly propositional, *thinkings* (in the narrow sense, to include acts of judging-that, wondering whether, supposing-that, and so on, but not including perceptual or quasi-perceptual states of seeing-that, or hearing-that, or imagining-that) which constitutively involve natural language. It is one thing to *notice* a car approaching, or to *imagine* the wind on my face, and another thing to *think to myself* that there is a car coming, or that it is windy today. And it is only the latter events which are language-involving, if introspection is to be believed – although, as we shall see in Chapters 7 and 8, it may well be the case that conscious perceivings and imaginings only achieve their status as conscious by virtue of being made available to a kind of thinking which *does* involve natural language.

I have suggested that the scope of the introspective thesis should be restricted to occurrent propositional thinkings. However, there may be a way in which it could be extended to include standing-state propositional attitudes as well. For we can draw a further distinction between beliefs and desires which are *dormant* and those which are *active*. Most of our beliefs and desires lie dormant for most of the time, continuing to exist, but without having any effect on current mental processes. My belief that Sheffield is in England is something that I retain throughout my life, but only rarely does it make any difference to what I think or do. Sometimes, however, those beliefs and desires become active, as when one forms new beliefs on the basis of old, or acts so as to satisfy a desire.

Now, it is highly plausible that one way in which a standing state can become active is by emerging as, or causing, an appropriately related episodic event. Thus, one way for my belief that roses are powerfully scented to become activated, is for it to cause me to think, in assertoric mode, ‘Roses are powerfully scented.’ And one way for my desire for fame to become activated, is for me to think to myself, ‘Would that I were famous!’; or something sufficiently similar. (These points will prove to be of some importance in Chapter 6.) Then if these latter episodic events involve natural language, as the argument from introspection suggests, it will follow that the standing states must also involve such language, at
least to this extent – that they are partly constituted by their disposition to
give rise to episodic events which involve natural language.

(It does not yet follow, of course, that one’s beliefs and desires are stored as, or in the form of, natural-language sentences. But considerations of
simplicity would certainly suggest it. If a large part of the way in which
beliefs and desires function actively in cognition is through natural lan-
guage, then it would certainly make sense that they should also be stored
in that form.)

It begins to seem plausible that a good deal of conscious human think-
ing, at least, might involve natural language. But this claim is still rela-
tively weak, for our purposes. For no modal conclusions are yet
warranted. That our conscious thinkings do occur in, or involve, natural
language does not yet show that they must, not even out of mere natural
necessity. It will be the task of Chapter 8 to try to establish this stronger
conclusion. But what does already follow from the considerations above –
at least if they stand unchallenged – is that our conscious thinkings do not
occur in Mentalese.

If my conscious thinkings take place in English, as I have claimed on
the basis of introspection that they do, then of course they do not take
place in any other natural language, let alone in a supposedly innate, uni-
versal, symbolic system such as Mentalese. So if the argument from intro-
spection is allowed to stand, we have already done enough to show that
Fodor and others are wrong to claim that all thoughts are expressed in
Mentalese. It may turn out to be the case that non-conscious episodic
thinkings occur in Mentalese, and it might remain possible, though
unlikely, that standing state thoughts should be stored in that form. But
conscious episodic thinkings would definitely seem, at this point, to
involve natural language. I shall return to consider Fodor’s possible
avenues of reply in section 2.5 below.

2.4 Objections and elucidations

The introspective thesis that we do much of our conscious thinking in
natural language sentences faces a number of intuitive objections. In this
section I shall address as many of them as I can, taken in no particular
order. Some of these objections will be returned to in much more detail in
later chapters.

(1) ‘A thought can surely occur to us in a flash, fully-formed and appar-
ently determinate in content, without there being time for any sentence
expressing that thought to be formulated, and without any such sentence
being introspectively accessible.’

But the reply to this is relatively easy, and is the one given by
Wittgenstein (1953, sections 318–20). Lightning-like thought can occur in the same way and for the same reason that it is possible to make a note of a thought in a few pencilled dashes – namely, what makes those dashes into the expression of a complete thought is my ability to expand them *into* such a thought, together with my present disposition to make use of them as if they *already* expressed such a thought. The phenomenon of lightning-like thought is one which is, in fact, equally characteristic of public speech or writing as of private thinking, and so cannot serve to undermine the claim that we think in natural-language sentences.

(2) ‘But surely we are often aware of a thought, or of entertaining a conscious thought, without any sentence figuring amongst the data available to introspection. For example, when choosing amongst a number of garments from a display I can surely judge that the right-hand one has the more attractive colour without the form of words, ‘The right-hand one has the nicest colour’, – or indeed *any* form of words – figuring in my consciousness. And this is confirmed in the experience-sampling data gathered by Hurlburt (1990 and 1993), since many subjects reported having wordless, purely propositional, thoughts on occasion.’

My response to this point is to grant that we often have thoughts which may not involve images, either of objects or of public-language sentences, but to deny that those thoughts are conscious ones. In part this reply will have to wait on the discussion of consciousness in later chapters. But very roughly, what I mean by a conscious thought is one that we are aware of ourselves having, when we have it, *non-inferentially* (in a way which is different from our awareness of the thoughts of other people). But what happens in cases of languageless (and imageless) apparently conscious thinking, in my view, is that the thought is, in fact, *not* so available. Rather, it is self-ascribed on the basis of a swift retrospective self-interpretation, much as it might be ascribed to a third party. I shall return to defend this way of looking at the issue in section 7.3, and then again more fully in section 8.3.

(3) ‘But what of various forms of tip-of-the-tongue phenomena? It is a familiar occurrence that people find themselves knowing (or at least believing that they know) what they want to say, but being unable, for the moment, to find quite the right words to express it – often with the sense that there is a word or phrase which exactly fits what they mean, if only they could recall it. Here, it seems, there is a thought which is both conscious and determinate prior to any natural-language sentence being found for its expression, which would suggest that the thought itself is independent of such language.’

But there are a variety of different kinds of phenomena involved here, in fact (as Wittgenstein again points out, 1953, section 335), none of which
2.4 Objections and elucidations

raises a real problem for the introspective thesis defended above. One sort of case is where there is, indeed, a determinate thought present, which is expressed, at least in part, in some other form of representation. Asked for directions to the City Hall I may find myself floundering, trying to express my determinate (imagistic) knowledge in terms of names of streets, and right and left turns. But since I have already allowed that some thinking – particularly spatial thinking – can involve visual and other images, there is no particular problem here.

In other sorts of case I believe it can plausibly be denied that there is any determinate thought in existence prior to its linguistic formulation. Rather, what happens is that I know (or believe) that there is thought there to be had, and am confident that I would recognise it if I found it. In fact, I am struggling to find – or to have – a thought of a certain sort, not struggling to express a thought which I have already entertained. The situation is rather like one in which, looking at some sort of visual puzzle or maze, I can be entirely confident that the puzzle does have a solution, and a solution of a certain sort, prior to the actual discovery of that solution. Evidence supporting this reading of the phenomena come from those not-infrequent cases where, on pressing further for the thought which I took to be there, I find that I was really just confused, and that there exists no thought of the kind that I wanted.

(4) 'Consider examples where one makes a choice between alternative words in expressing a thought. Does this not show that the thought itself must already exist in consciousness independently of the words used to express it?'

In fact not. And once again, there are really two sorts of case here. One of them parallels the phenomenon just discussed. Sometimes choosing between alternative words is choosing between alternative, slightly different, thoughts – one of which may be more appropriate to the context, or each of which may be appropriate in different ways.

The other sort of case just requires – what I have anyway allowed – that some of our non-conscious thoughts may fail to involve natural language. In cases where I choose between alternative ways of expressing a thought, I am prepared to allow that there is, sometimes, a thought already there to be expressed; but I deny that it is a conscious thought. Similar points apply to the processes which constrain and guide the way in which we fit together words and phrases to form sentences – either aloud or in imagination. I can allow that such processes are thoughtful, but deny that the thoughts in question are conscious ones. My hypothesis – to be pursued in later chapters – is that thoughts only become conscious when they emerge into episodic imagings, either of public sentences alone, or of mixtures of sentences and of objects or places.
(5) 'But many sentences of natural language are ambiguous. To take a hackneyed example, the sentence, “John is at the bank”, can mean either “bank-of-a-river” or “banking-institution”. Yet my thoughts are, surely, not similarly ambiguous. When I think, “I shall go to the bank in the lunch-hour”, there is normally no indeterminacy in my meaning. In which case my thoughts cannot be identified with the sentences used to express them.’

I have two replies. One is that it may not be the sentence, as such, which constitutes a given thought, according to the introspective thesis; it is, rather, the sentence in use. An occurrent thought will consist in an imaged sentence taken together with my dispositions to reason and act on that sentence in one set of ways rather than another. So it is not just the imaged sentence, ‘I shall go to the bank’, which constitutes my decision to attend a banking-institution; it is that together with a background of dispositions and abilities. (On the idea of ‘the background’ to thought, see Searle, 1983, ch. 5.) And it may well be the case that these in turn do not similarly involve natural language. But then neither are they conscious. So the thesis that it is our conscious workings which consist in deployments of imaged natural language sentences can remain intact.

The second reply is that an imaged sentence is not, in any case, a merely phonological representation; rather, the content is represented in the image too. The imaged sentence, ‘I shall go to the bank’, carries with it one or other possible interpretation, just as the heard sentence would do. When we hear people speak, we do not just hear phonology, and then have to figure out an interpretation. Rather, what we hear is already interpreted. If someone says to me, ‘I shall go to the bank’, then I will immediately hear this as meaning one thing or another (depending on the context) – though I may have to think again if the meaning I hear is sufficiently incongruous. So it is, too, with inner speech. Ambiguous sentences always carry with them some particular interpretation when imaged. Putting the point in terms of the reply of the previous paragraph: the background dispositions which constitute my meaning the sentence in one way or the other may be realised in some sort of representational structure (of Chomsky’s logical form, or LF, as it might be) which is already attached to the phonological representation. So the image is actually an interpreted object – it is not ambiguous.

(6) ‘But do not creative thinkers like Einstein often insist that their thinking is done without words, but rather through fleeting images and feelings which seem to bear no relation to words? (see Ghiselin, 1952). And do not the solutions to problems often just come to us, already formulated? In such cases one must have been thinking, in order to have reached a solution; but the thinking was not done in introspectively accessible words.’
These points are good ones. But to say that creative thinking is done wordlessly may only mean that it is done non-consciously. Not for nothing have poets traditionally prayed to the Muses for inspiration; for we often have no idea where our genuinely novel ideas come from, nor is there much that we can do intentionally in order to get them. Sometimes a relaxing environment can help – a hot bath, a daydream, or a good night’s sleep. But when ideas do come, they seem to us to come of their own accord, often with no discernible history.

As has often been remarked, indeed, it may be that intelligence divides into two separate components – a creative, non-discursive, non-conscious idea-generator on the one hand, and a judgemental, logical, discursive, conscious planner and idea-assessor on the other. In which case, it may very well be the case that when engaged in creative thinking, all that will pass through my consciousness may be irrelevant and fleeting images and feelings. For the thinking itself is non-conscious. This is perfectly consistent with the claim to be defended in this book, that conscious thinking involves language. (It is also consistent with the claim, note, that the non-conscious creative thinking is also language-involving. It is hard to see any reason why language could only figure in thought in conscious mode. I shall return to this point in section 8.7.)

(7) ‘But what are you to say about conversational implicatures, and about pragmatics? Consider metaphor, for example. Suppose that an irritated teacher remarks of a pupil, “That fat slug is just waiting for his lunch”. Here what is meant – the thought expressed – is something like: that the pupil is fat and slow and indolent, and thinks about food instead of working. But what is said is something quite different – something which either fails to refer at all, since there is no appropriately salient slug, or which says, absurdly, that the pupil in question is a slug. Then, since there is a mismatch between what is thought and what is literally said, must not the two be wholly separate? The thought might be formulated in one way – in Mentalese, as it may be – but then given its public expression quite differently.’

This way of looking at the matter is by no means mandatory, however. The fact that a particular sentence-token of natural language only obtains its intended content, in a particular context, when taken together with a good many surrounding beliefs and intentions, does not mean that the occurrence of that content is independent of the sentence-token in question. Rather, we can say that this token occurrence of that thought is inseparably bound up with its (non-literal) linguistic expression, but also depends for its existence on a variety of other thoughts which make it the case that the latter is constitutive of the former.
2. Which language do we think with?

2.5 Fallible introspection and Fodor

Recall from section 1.8 that I am accepting, with Fodor, that thoughts are sentences which are tokened, stored, and processed in characteristic ways. And recall, too, that Fodor's claim is that all thought is conducted in sentences of Mentalese. But now the introspection-based argument developed above maintains that our conscious thoughts (at least when episodic and properly propositional) are sentences of natural language. So there is an inconsistency: one or other of this trio of claims must be false. How should Fodor respond?

One way in which he could respond, would be by denying that introspection is reliable on such matters. If Fodor is to maintain that all thoughts, including conscious episodic ones, are expressed in Mentalese, in the face of an argument from introspection to the contrary, then he could try to find some way of rejecting the deliverances of introspection. This need not necessarily be as difficult as it sounds. For hardly anyone, nowadays, maintains that introspection is strictly infallible, in true Cartesian tradition. Most hold that there is scope for error in our awareness of our own mental states, just as there is in our awareness of objects outside us. So if Fodor claims that introspection misleads us into thinking that our conscious thoughts involve natural language when they do not, this cannot plausibly be refuted by an appeal to introspective infallibility.

However, most people today also think that introspection, like external perception, is generally reliable. True enough, there are some people who deny that mental states really exist at all (see Churchland, 1981). In which case introspective awareness could hardly be claimed to deliver us with truths about our mental lives. But almost everyone who believes that there are truths about mental states available to be had – whether realists, like myself and Fodor, or interpretationalists, like Davidson and Dummett – maintains that introspection will generally be reliable in getting us those truths. If I think that I feel pain, then I generally do; if I think that I am thirsty, then I generally am; if I think that I am fantasising, or reasoning, or deciding, then that is generally what I am doing; and so on. It is, therefore, not enough for Fodor merely to appeal to the possibility of introspective error. Some reason needs to be given why the error should be systematic and persistent. For if Fodor is correct, then we are all of us systematically misled into believing, on the basis of introspection, that we think in natural language.

But Fodor does not need to deny the phenomenon of inner speech as such, in fact, nor the reliability of introspection in relation to it. That is, he need not reject the introspective datum that images of natural language sentences figure prominently in consciousness when we think. But what
he must deny is that such sentences constitute our thoughts. My claim is that when, in the course of trying to solve a practical problem, I am aware of imagining the sentence, ‘I will get it by climbing on the box’, then what I am aware of is the thought. My claim is that it is the imaged sentence itself, in context, which is the occurrent intention, and which then sets in train the motor routines which cause my body to climb up on the box. And this is the way in which I would naturally report my introspection – I was aware of deciding to climb on the box (and not just aware of something caused by my decision), and it is because I so decided that I did it. Our common-sense construal of the introspective datum, is that the imaged sentences which we are aware of occupy the causal roles in our cognition distinctive of thinking.

Fodor must claim that both I, and common sense, are mistaken. What I am aware of here is not the occurrent intention itself (which is, he will say, expressed in Mentalese), but some mere concomitant of it. My dispute with Fodor can thus concern the causal roles of imaged natural-language sentences, rather than their existence. The dispute can be about whether natural-language sentences occupy the causal roles within our cognition of thinkings, decidings, and wonderings whether; not about whether such sentences figure in cognition at all. My claim is that it is the imaged sentence itself – ‘I shall get it by climbing on the box’ – which occupies the causal role of the thought (the thought, namely, that I will get it by climbing on the box). Fodor’s claim must rather be that the thought itself, expressed in Mentalese, somehow causes the imaging of the natural-language sentence (as well as causing me to climb on the box).

If Fodor’s views on the nature of thinking are to be vindicated, then what he needs is either a full-blown theory of consciousness (a theory of introspective self-awareness), or at least a plausible account of the role of inner verbalisation within our cognition (if that role is not the one distinctive of thinking). He therefore needs at least enough of an account of introspection which will not only be plausible and convincing in its own right, but will at the same time explain how we come to be under the systematic illusion that our conscious thoughts involve natural language. This is not something that he actually provides anywhere in his writings. So we shall need to consider what can be done on his behalf.

We shall need to canvass a variety of theories of consciousness, and a variety of accounts of the causal role of inner speech, to see whether a thesis of systematic illusion can be made out. If the most plausible of these theories turns out to have the consequence that Fodor requires, then he will face no problem from the introspective argument presented in this chapter. He can reply that our best theory of consciousness, and/or our best account of inner speech, implies that we shall all easily become
deluded on this matter. So here is one reason why the question of the extent to which thought involves natural language should lead us to investigate theories of consciousness, as we shall do in Chapters 5 to 7. (Another reason will be given shortly, in section 2.7.)

Suppose, however, (and as I shall argue, in fact) that our best available theory of consciousness, and our best account of the function of inner speech, do not provide us with any grounds for believing that we might be systematically deluded about the nature of conscious thinking. Would this mean that Fodor’s position – that all thinking is conducted in Mentalese – could not be rationally believed? Not necessarily. That would depend upon the strength of his arguments in support of his thesis. For suppose that those arguments were very powerful. Then provided that there were some theory of consciousness, or some theory of the function of inner speech, which would support the idea of systematic introspective illusion, then the most rational course might be to endorse that theory, even though it is, on other grounds, somewhat weaker. Indeed, even if none of the available theories were to support the idea of systematic illusion, this need not mean that Fodor’s position could not be rationally believed. For if his arguments are strong enough, the best course might be to insist that introspection must be mistaken on this issue, and to maintain, in consequence, that no one has yet managed to propose an acceptable theory of introspective awareness, or an acceptable account of the role of inner speech.

This means that we shall have to spend some time considering the strength of Fodor’s arguments for the thesis that all thought occurs in Mentalese. This task will be begun in the next section, and will occupy us through to Chapter 4. In the present chapter I shall confine myself to three of Fodor’s more explicit arguments, as developed in his 1978 (see also his 1987, Appendix).

2.6 Individuating propositional attitudes

The first of Fodor’s arguments that I shall consider here, proceeds by attacking the contrary claim that beliefs and desires are relations to sentences in the natural language which is native to the believer or desirer in question. The argument is, that claiming propositional attitudes to be relations to natural language sentences will slice them too thin. That is, if beliefs and desires are individuated in terms of the natural-language sentences which are claimed to constitute them, then the resulting identity conditions will lead us to distinguish beliefs and desires which are, in reality, the same.

It will then turn out, for example, that the belief that the dog bit the man
is distinct from the belief *that the man was bitten by the dog*, because the natural-language sentences, ‘The dog bit the man’, and, ‘The man was bitten by the dog’, are different from one another. Whereas we would surely want to claim that those beliefs were one and the same. Similarly, it will turn out that Oedipus’ desire to marry Jocasta is distinct from his desire *that Jocasta be his wife*, since the sentences, ‘I am married to Jocasta’, and, ‘Jocasta is my wife’, are different. But again, we would surely want to insist that these desires are identical.

The principle of propositional-attitude individuation appealed to here derives from Frege (1892), and is often known as *the intuitive criterion of difference*. On this account, thoughts are distinct in content if, and only if, it is possible for a thinker to take differing epistemic attitudes towards them at one and the same time. So the belief *that Venus has set* is distinct from the belief *that the evening star has set*, since it is possible for someone, not knowing that Venus and the evening star are one and the same, to believe that Venus has set while doubting that the evening star has. By the same token, Oedipus’ desire *to marry Jocasta* is distinct from any desire *to marry his mother*, since it is plainly possible to have the one desire but not the other – as, in fact, he did. In contrast, it does not seem possible that someone should believe *that the dog bit the man* while doubting whether *the man was bitten by the dog*, except out of momentary confusion or madness. Nor does it seem intelligible that Oedipus should want to marry Jocasta while having an aversion to the thought that Jocasta should be his wife.

It is worth noting that while the intuitive criterion of difference slices thicker than natural-language sentences, it still cuts propositional attitudes pretty finely – in particular, more finely than logical equivalence. Thus, by the criterion, the belief *that there are 9 planets* will come out as distinct from the belief *that the number of planets is two larger than the smallest prime number larger than 5*. This is so for two distinct reasons. First, it seems plain that someone could believe the first while being inclined to doubt the second, because of errors in calculating the result. Second, the latter belief contains concepts – for example, the concept *prime number* – not contained in the former. So someone could believe the former while failing to believe the latter because they lacked the concept of a prime number.

Fodor’s claim, then, is that once we have accepted that propositional attitudes are relations to sentences, the only theory consistent with the principle of individuation contained in the intuitive criterion of difference is that the sentences in question are sentences of Mentalese. The hypothesis can then be, that the sentences, ‘The dog bit the man’, and, ‘The man was bitten by the dog’, will both translate into the very same sentence of
Which language do we think with?

Mentalese, as will each of the sentences, ‘I am married to Jocasta’, and, ‘Jocasta is my wife.’ So the reason why someone must assent to each of these pairs of sentences if they assent to either, is that they are both equally acceptable public translations of the Mentalese sentence which really constitutes their propositional attitude.

In reply to this argument, it is only necessary to introduce the distinction between core and dispositional propositional attitudes, deployed by Field, 1978, following Dennett, 1975. A core belief, for example, is one which has been explicitly computed and stored in memory – in the form of some sort of sentence, if the sentential account of propositional attitudes is correct. A dispositional belief, in contrast, will be any obvious consequence of a core belief, apt to be explicitly computed as soon as circumstances demand. Now, before I put this distinction to work in answering Fodor’s argument, I shall show how it is independently motivated – that is, I shall show how the distinction is one which needs to be drawn anyway, whether we believe that propositional attitudes are relations to sentences of natural language, or of Mentalese.

The distinction between core and dispositional beliefs is required to explain how we may correctly attribute an unlimited set of beliefs to creatures, such as ourselves, with finite cognitive space. I surely say something true of you when I say that you believe that 61 is larger than 60; that 62 is larger than 61; that 63 is larger than 62; that 64 is larger than 63; ... that 73,000,001 is larger than 73,000,000; and so on without limit. Equally (to use Dennett’s more colourful example), I surely say something true of you when I say that you believe that zebras in the wild do not wear overcoats; that they do not hold barbecues; that they do not read Shakespeare; and so on, again without limit. But it is highly implausible to suggest that each of these beliefs corresponds to a distinct state of your brain (a stored sentence). For there are, to put it mildly, too many of them. Rather, each is an obvious consequence of some body of explicitly stored beliefs, in such a way that you will, for example, immediately answer ‘Yes’ if asked whether 73,000,001 is larger than 73,000,000. And note that something like this will need to be said whether beliefs are stored in natural language, or of Mentalese, or some other system of representation altogether.

Now, with the distinction between core and dispositional beliefs in place, we can say that the belief that $P$ is none other than (is the very same as) the belief that $Q$ whenever the sentence ‘$P$’ is obviously equivalent to the sentence ‘$Q$’. For then it will not matter, for purposes of psychological explanation, which sentence we use to characterise the subject’s belief. For if either belief is core, then the other will be dispositional. No matter whether it is the sentence ‘$P$’ or the sentence ‘$Q$’ which is stored in memory, the person will be disposed immediately to compute the other
sentence if required. So if they assent to the question, 'Do you believe that P?', then they will also assent to the question, 'Do you believe that Q?', and vice versa.

We can now reply to Fodor's argument against the thesis that beliefs are relations to sentences of natural language, by saying that, more strictly, beliefs are relations to sets of obviously equivalent natural-language sentences; and that a core belief will consist in the storage of any one of a set of obviously equivalent natural-language sentences. Indeed, it is an additional virtue of the natural-language account that it can provide a substantive explanation of the condition for thought-identity captured in the intuitive criterion of difference. It is because it makes no apparent psychological difference whether it is the sentence 'P' or the sentence 'Q' which is stored, if 'P' and 'Q' are obviously equivalent, that it will make no difference whether the thinker is described as believing that P or as believing that Q.

('Obvious equivalence to whom?', you might ask. Obvious to the subject to whom the beliefs are being attributed, in the first instance. It is because the subject in question will immediately assent to 'P' if they are already inclined to assent to 'Q', and will similarly assent to 'Q' if they are already inclined to assent to 'P', that it makes no significant difference whether we describe them, in either case, as believing that P or as believing that Q. But it is, on the present hypothesis, a presupposition of our practice of ascribing beliefs in accordance with the intuitive criterion of difference, that such patterns of obviousness should be widely shared. I rely on the fact that I find 'P' and 'Q' to be obviously equivalent when I judge that these two sentences should be counted as expressing the same belief. And generally such judgements are borne out in psychological practice.)

2.7 Animals and infants

Another of Fodor's major arguments for the thesis that the language in which we think is Mentalese, is that animals and pre-linguistic human infants have thoughts - beliefs, desires, intentions, and the rest - but no natural language. In which case the sentences constitutive of their thoughts cannot be natural-language ones. Now, it is not very plausible to deny the premise of this argument. While some have denied that animals and infants have beliefs and desires, their grounds for doing so have either been weak, or have depended upon some anti-realist conception of the mind. At any rate, their position is not one that I propose to take seriously in this book. (For further discussion, see my 1992c, ch. 6.) It is much more plausible to respond by denying that the conclusion follows from
the premise – not even non-demonstratively, by means of an inference to the best explanation.

In the first place, the most that follows from the claim that non-human animals have propositional attitudes, of course, is that they entertain their thoughts in some form of Mentalese, whose sentences perhaps have the same meanings as some of the simpler sentences of human natural language. There seems no particular reason to believe, as yet, that human thoughts will employ the very same medium of representation, especially given the manifest differences that exist between human and animal modes of cognition.

The case of pre-linguistic human infants might seem more problematic, however. For if we concede that their thoughts occur in Mentalese, then the simplest hypothesis would appear to be that all human thoughts are similarly expressed. For it is simpler to suppose that human cognition employs just one system of representation rather than two. Not that this conclusion is by any means forced on us. For there remains the counter-suggestion put forward by Field (1978), that natural-language propositional attitudes get grafted onto a more primitive system of Mentalese employed by pre-linguistic children. Indeed, when put together with the evidence from introspection, presented earlier in this chapter, this might seem to be the most plausible proposal overall.

Field’s counter-suggestion can be strengthened still further, however – even to the point of being able to stand independently of the evidence of introspection – if we can identify some major functional difference between the thoughts of adults and those of pre-linguistic infants and animals. For otherwise, why should it be supposed that there are two quite different systems of representation in operation, especially within the cognition of a single organism? This will occupy us extensively in the latter half of this book, where I shall argue that the difference in question corresponds to the distinction between conscious and non-conscious propositional attitudes. In fact this argument of Fodor’s will only finally be answered in section 7.8, at the conclusion of my discussion of the nature of consciousness. I shall then go on to claim that conscious thinkings are relations to natural-language sentences, whereas non-conscious ones are – or may be – relations to sentences of Mentalese.

(In fact I shall remain agnostic on the question whether the non-conscious thoughts of animals and pre-linguistic infants are best understood on the sentential model, or whether they should be thought of as imagistic, or whether, indeed, they might be better approached from the standpoint of so-called connectionist cognitive architectures. I shall also remain agnostic about whether or not the central-process propositional attitudes of animals and pre-linguistic infants employ the same systems of repre-
2.8 Language-learning and sub-personal thought

Yet another of Fodor's main arguments for the thesis that propositional attitudes are relations to sentences of Mentalese, turns on the question of how natural languages are learned. I shall first present and criticise this argument, before introducing some relevant considerations to do with the modularity of the mind.

The argument

Fodor claims that the only theory of learning which we have, is that learning always involves an inference to the best explanation. On this view, learning, of whatever kind, always involves the stages familiar to students of scientific method, of data-collection, hypothesis formation, testing, and confirmation. So when an infant learns the meaning of the word 'cat', for example, it must first gather some data about adult usage, then formulate a hypothesis, such as that 'cat' means 'furry animal', and then confirm or refute (and subsequently modify) that hypothesis in the light of further observation.

All of this presupposes, of course, that the infant has some symbolic system in which it can describe and record the initial data, and in which it can express and modify its hypotheses. It is hard to see how anything could serve, here, except an articulate language – indeed, a language with the full expressive power of human natural language. But, plainly, this symbolic system cannot be a natural language, on pain of vicious circularity. Since the symbolic system in question is the medium of representation through which natural languages are supposed to be learned, it must be a universal, presumably innate, language of thought – that is to say, Mentalese.

There is an obvious initial reply to this argument, which appears to have at least some independent motivation. It is, that knowledge of natural language is not so much a matter of knowledge that – that is, of propositional knowledge – as of knowledge how, the kind of knowledge which is involved in possessing a skill, or practical capacity. On this view, learning your first natural language is not like learning a body of facts, such as quantum physics, but is more like acquiring a practical skill, such as the ability to ride a bicycle. For we do say that the child is learning how to speak, and knowledge of the meaning of a word is quite naturally assimilated to an ability to use it correctly. And presumably no one thinks
that, in order to learn to ride a bicycle, the child must formulate and test hypotheses about such matters as the correct angle at which to lean into a corner that is taken with a given angular velocity. So why should we think that the child learning its first language must be formulating hypotheses about the meanings of words, either?

Fodor has available to him a twofold rejoinder. First, he can point out that if knowledge of a natural language is a practical capacity, it must be a special sort of capacity, whose categorical basis in the brain somehow reflects the semantic and syntactic structures of the language in question. For no one who is not a behaviourist thinks that learning a natural language is a mere matter of behavioural training, or of developing the right conditioned responses. This is because of the unlimited creativity of language. Anyone who masters a natural language has the capacity to understand, to use, and to recognise the well-formedness of unlimitedly many distinct sentences. The basis of their capacity must, then, at least be structured into distinct recombinable components, rather as sentences are composed of distinct recombinable words. It is another matter, however, to claim that the basis of the capacity to use a natural language is properly characterised in terms of propositional knowledge, as Fodor does. In fact, we can continue to maintain against him that knowledge of a natural language is a practical capacity, albeit one whose categorical basis is appropriately structured. (For further discussion of this issue, see my 1992a, ch. 6.)

The second rejoinder available to Fodor, and the one he actually pursues (at least implicitly), is more radical. It is to claim that knowledge how is, at a deeper level, to be assimilated to knowledge that. For he claims that all cognitive processing, at all levels of cognition, involves computations on language-like representations. His argument for this is an argument from the presuppositions of much cognitive science, of which David Marr’s 1982 might serve as a suitable example. Marr and others in the study of sub-personal systems of vision, language-acquisition, face recognition, and so on, postulate rich hierarchies of computation and inference between the incoming data and the output of those systems. So it may turn out, it appears, that even learning how to ride a bicycle is to be explained in terms of postulated processes of hypothesis formation and testing in some non-conscious, sub-personal, symbolic system. In which case, the claim that language-learning is a matter of learning how will have gained us nothing against Fodor.

How convincing should we find Fodor’s second rejoinder? That depends on how plausible we find the alternative – connectionist – approach to cognitive science. Those working within this alternative tradition have attempted to construct models of cognitive processes
(including various forms of learning) which involve, not computations on sentence-like representations, but rather distributed activations between connected neuron-like units (see, for example, Rumelhart and McClelland, 1986; see also Clark, 1989).

Now, it is not necessary that we should commit ourselves wholeheartedly to connectionist approaches to cognition in order to be able to reply to Fodor's second rejoinder. It is possible to be convinced, as I am, by Fodor's arguments in support of more traditional representational theories of mind (see his 1987, Appendix, and Fodor and Pylyshyn, 1988), but only in so far as they apply to central processes of thinking and reasoning. We could therefore continue to insist that beliefs, desires, and other propositional attitudes involve relations to sentences, while allowing that connectionist approaches may be able to account for many sub-personal cognitive processes, including those involved in language-acquisition. And even if language-acquisition, in general, is not adequately accounted for in connectionist terms (as Steven Pinker argues, 1988), it may be that vocabulary acquisition is different. It may well be that those aspects of language which are very definitely learned and not innate, such as vocabulary, can be adequately accounted for using connectionist models, leaving rule-based systems, of the sort postulated by Chomsky's innate universal grammar, to explain the rest. At any rate, this possibility has not yet been ruled out.

There is available, however, a much simpler and swifter response to Fodor's argument from language-learning. For we could allow that language-learning requires thought, but deny that the thoughts in question are conscious ones. Since the thesis to be pursued in this book is only that conscious thinking involves language, I need not be troubled by an argument whose only real upshot is that conscious thinking presupposes various forms of non-conscious thought. For such a claim is independently very plausible.

Yet in order for this reply to be convincing, enough needs to be said about the nature of consciousness for it to be apparent that the kinds of thought entered into by pre-linguistic, language-learning, children would be of the non-conscious variety. That, obviously, must be a task for later chapters. But, very roughly, I shall claim that conscious thoughts are those that are reflexively available for the subject to think about in turn – see sections 7.1 and 7.3 below. What makes an occurrent thought of mine conscious, on this account, is that it occurs in such a way that I can then go on to think, not only about what I have just thought, but also about the fact that I have just thought it. And I do believe that it is immensely implausible that pre-linguistic children should be capable of such thinking – see section 7.8.
Modularity of mind

The argument from language-learning appears less than convincing, then. We can, in addition, turn Fodor against himself, appealing to the thesis of modular mental organisation which he has himself defended elsewhere (see his 1983 and 1989). Fodor argues that the mind is organised into input and output systems – or modules – on the one hand, and central processes on the other. Modules would include vision, audition, language processing, and motor control. Central processes would include belief, desire, and practical reasoning. The distinctive features of mental modules are that they are isolated, innately specified, and fast. I shall say a few words about each of these features in turn.

Mental modules are held to be isolated from the rest of cognition, being largely impervious to changes in central belief. Thus, for example, people continue to be subject to the Mueller-Lyer illusion, continuing to see the arrow-tailed line as longer than the arrow-headed one, even when they know that the lines are of equal length. At some level of perception what you see will depend upon background belief, of course – thus, whether you see something as a car, or as a cup, will depend upon a whole battery of beliefs about cars and cups. But whether you see a particular three-dimensional solid shape against an immobile background is not similarly a function of belief, and will arguably remain the same no matter how your beliefs may change.

Mental modules have this isolation because they are dedicated processors, assigned a specialised role in cognition, much of whose operation is innately specified, or ‘hard wired’. As a result of this specialisation, mental modules are extremely fast in relation to central processes. For example, subjects are capable of shadowing speech with a latency of 250 microseconds, thus allowing about 125 microseconds – or one eighth of a second – for the analysis of speech input. Since the speed of neural transmission is only 10 microseconds, this means that analysis of a heard sentence (or, indeed, of a visual scene) takes place within the space of 12 neural firings. This is, by any measure, fast.

Now, given the isolation of input systems from the rest of cognition, there is no particular reason to think that they function in the same way as central processes of thought, or that they employ the same system of representations. They merely have to deliver their results to central processes, or, in the case of output systems, take input from those processes. Moreover, given the speed of input and output systems in relation to conscious thought, there is some reason to think that mental modules do not operate in the same way, or employ the same system of representation.

These points seem to me to undermine the unity of the language of
thought, leaving us free to maintain that propositional attitudes are relations to natural-language sentences, whereas input and output modules operate in some other medium – sentences of Mentalese, or connectionist nets, as it may be. So, granted that the mechanisms for language-acquisition may employ some sub-personal system of representation, it does not follow that personal thoughts must also employ that same system of representation. Indeed, as we saw earlier in this chapter, there is some reason to think that they do not.

It is worth noting, moreover, that although the states attributed by cognitive scientists to sub-personal systems dealing with vision or language can sound like propositional attitudes – for example, formulating a hypothesis, or addressing a question to a sub-processor – they do not have the essential functional characteristics of genuine propositional attitudes. In particular, these states play no role in anything genuinely analogous to practical reasoning. Thus, real (central process) hypotheses function in the fixation of beliefs, which in turn interact with desires in characteristic patterns of inference to produce intentions, which interact further with other beliefs and perceptions so as to produce actions. In sub-personal cognitive systems, in contrast, it does not seem plausible that there is any real place for desire, for intention, or for action. So the moral is, again, that there is no particular reason to think that sub-personal modules will employ representations of the same general kind as central thinking processes.

Suppose that the thesis of modular mental organisation is correct; and, in particular, that there is a dedicated language module which is isolated from central processes of thought and belief. This might then seem to cause a problem for the claim to be defended in this book, that human conscious thought involves natural language. For how can natural language be implicated in thought, if the language module is isolated from thought? The answer is that isolation need not be reciprocal. While the functioning of the language module may be largely isolated from changes in background belief, it may be that thought and belief can have access to some of the structures and information inherent in the language module.

This point will be expanded upon in Chapter 8, where I shall postulate the mechanisms subserving human consciousness. Here I merely note that just such an asymmetric relation seems to obtain between the visual system and central processes involving imagination, for example. There is evidence that while the visual module may be isolated (up to a certain level of visual output) from central belief, visual imagination itself accesses and employs some of the cognitive resources involved in visual perception. Thus, a task requiring visual imagery, such as counting the right-hand corners in an imaged F, will be interfered with if the responses have to be given visually (for example, by pointing to ‘Yes’ and ‘No’
Which language do we think with?

buttons), but not if they may be spoken. In contrast, a task requiring auditory imagination, such as counting the nouns in a recently heard sentence, will be interfered with by a spoken, but not by a visual response (see Fodor, 1975, and Tye, 1991). It seems that imagination requires, and has access to, some of the cognitive resources of the appropriate sense modality.

Summary

In this chapter I have raised the question whether it is natural language, or rather Mentalese, which is constitutively involved in our thinking. I began by considering the evidence from developmental psychology, and from aphasia, which was found to be equivocal. I then presented an intuitive, introspection-based, argument for the view that human conscious thinking constitutively involves natural language. But this conclusion has, at best, the status of a bare matter of fact, with no necessity of any sort attached. And in any case, it could easily be overturned by a sufficiently powerful argument for the thesis that all thinking occurs in Mentalese. I have also discussed three of Fodor’s arguments for this latter thesis (the arguments from propositional-attitude individuation, from the thoughts of animals and infants, and from language-learning) which were shown to be definitely unsound. But there remain a variety of further arguments to be discussed over the next two chapters, many of which have to do, roughly speaking, with questions of semantics.