4

Modularity and Flexibility: The First Steps

Is there any special puzzle or problem about developing an acceptable form of massively modular conception of the human mind, given the relatively weak construal of 'module' that we have adopted? Why would anyone think that the mind *shouldn't* be modular, in that weak sense? Some have argued that the mind *cannot* be massively modular, of course (Fodor, 2000); but they have employed a much more demanding notion of modularity, according to which modules have to be encapsulated in their processing. Once we retreat to the weaker notion of modularity articulated and defended in Chapter 1, then it is far from obvious that these arguments should retain their force. So is there any particular challenge remaining for massive modularity theorists to answer?

In Section 1 I shall articulate a number of such challenges. Thereafter (in this chapter and the chapters following) I shall discuss how those challenges should best be met. I shall be arguing in the present chapter that mental rehearsal of action (especially speech action) plays a crucial role in linking together and combining the outputs of some other central / conceptual modules, and in facilitating cycles of language-dependent activity, in so-called 'inner speech'. That role also makes possible a new *form* of language-based thinking and reasoning, I shall argue, realized in the operations of an underlying set of conceptual modules. If these accounts can be made to work, then the result should be highly attractive to massive modularists. For it is widely agreed that language is itself a module consisting of further sub-modules, as we saw in Chapter 3. In which case it might be the addition of a language module to the mix of modules that make up the human mind that is responsible for much of the latter's characteristic flexibility.

I believe, then—and will argue herein—that natural language has an important role to play in the distinctive flexibility of the massively modular human mind. But the thesis is, of course, an empirical one. And it should be acknowledged that most of the evidence required to support it just hasn't

been looked for or collected. (Some alleged evidence will be considered in Section 4.) But even if my conclusion were merely that it is causally (as opposed to logically) *possible* that language should play such a role, this would still be a result of significant interest and importance. For most of the people who reject massively modular models of the human mind do so because they can't see how minds with the flexibility of ours could *possibly* be modular in organization. I aim to show them how. One goal of this chapter, then, is to answer a 'How possibly?' question: how could the human mind *possibly* be composed of a massive array of modules? But I shall also hope to show that the proposed account of the role of language in cognition is, moreover, a *plausible* one, worthy of both further theoretical development and experimental investigation.

1 The Challenges

Recall that the thesis of massive modularity articulated and defended in Chapter 1 has the following form. The mind consists of a great many distinct processing systems (roughly one for each evolutionarily stable function or capacity, plus many others constructed through learning). The properties of these systems can vary independently of one another, their operations can be separately affected by other factors, and many of them can be damaged or destroyed without completely undermining the functionality of the whole arrangement.

We should also expect that there will be a good deal of variation in the degree of connectedness amongst modules. (See Figure 1.3.) For which other systems a module can receive input from, and where it will make its outputs available, will be a function of the processing task undertaken by the module in question, as well as the processing tasks undertaken by the others with which it is connected. But it almost certainly isn't the case that every module will be connected up with every other, since such connections will be costly to build and maintain, and since the addition of each such connection will make processing significantly less frugal. Roughly, there should only be connections where there really *need* to be connections (Coward, 2001).

As we also stressed in Chapter 1, the processing undertaken by mental modules will need to be frugal in terms of time and resources. They will all of them thus be encapsulated in the wide-scope sense distinguished in Chapter 1, although many might also be encapsulated in the stronger narrow-scope sense. (This means that *all* modules should have internal processes that require them to consult only a small subset of the total information available in the mind in the course of their processing, but some will also be restricted in the *kinds*

of information that they can look at—i.e. they will have a module-specific data-base.) And it will be very rare indeed that one module should have any access to the internal processes of another (as opposed to the outputs of one or more of the sub-modules contained within that other). Rather, other modules will at most have access to the results of that processing. So in addition to being wide-scope encapsulated, the internal processing of all modules should be *inaccessible* to most (if not all) other systems.[1]

FN:1

1.1 Massive Modularity and Common Sense

This form of massive modularity hypothesis predicts, then, that the mind decomposes into far *more* components than would generally be recognized, either by common-sense psychology or by regular (non-evolutionary) cognitive psychology. So part of the task before us is to articulate an architecture that can make sense of this. We need to say enough about the various modules and their mode of connectivity, either to explain how the common-sense picture can nevertheless be broadly correct in its outline framework; or to explain how the common-sense account can be so successful while being radically wrong.

This can be considered our first challenge. But meeting it is a straightforward matter. For I have suggested in Chapter 2 that the massive modularity thesis is best developed within the framework of a perception / belief / desire / planning / motor-control psychology. And then the basic architecture postulated by common sense will actually be correct. (Percepts give rise to beliefs and serve to inform practical reasoning; beliefs and desires interact in practical reasoning to create intentions and actions; percepts guide the execution of those actions.) Common sense's only failing will be that it doesn't postulate *enough* perceptual mechanisms, nor nearly *enough* mechanisms for producing new beliefs, new desires, and new actions. Compare Figures 4.1 and 4.2, in this regard. One difference between them is that in Figure 4.2 the visual system has been bifurcated, in accordance with the 'two visual systems' hypothesis of Milner and Goodale (1995). Another is that in place of some sort of unified theoretical reasoning system, there are now *multiple* reasoning systems for generating beliefs (and desires) in different domains. In addition, there are now multiple (and competing) decision-making systems,[2] and multiple motor-control systems.

FN:2

[1] Recall that inaccessibility and encapsulation are matters of degree. The conclusion of Chapter 1 was that the mind should be constructed out of a great many modular systems that have internal processes that are *largely* inaccessible and (wide-scope) encapsulated.

[2] Recall from Chapter 2.8, however, that the precursor architecture does allow for a sort of *virtual* unified practical-reason system, utilizing mental rehearsal and somasensory monitoring. This isn't represented in Figure 4.2. See Figure 2.8.

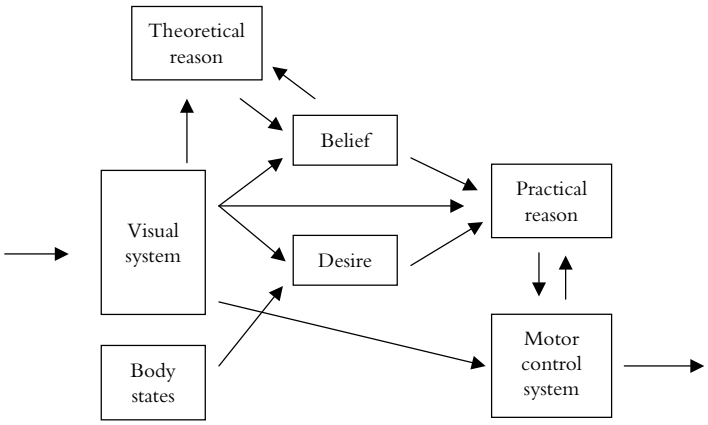

Figure 4.1. The common-sense mind

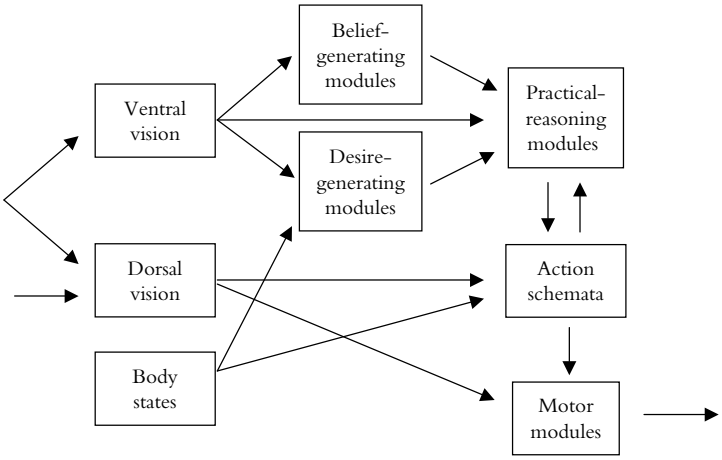

Figure 4.2. Multiple modules and dual visual systems

There is a different sort of objection to massive modularity that can be raised from the perspective of common sense, however. This is that it doesn't *feel* to us, on the inside, as if our minds were composed of massively many modules acting both sequentially and in parallel. On the contrary, we have the impression that the mind is diaphanous, or transparent to itself, with everything that happens within it occurring in a single unified arena containing conscious experience, conscious thought, and conscious decision-making. This is the intuition to which Locke (1690) gave voice when he wrote that there could be nothing within the mind that the mind itself was unaware of. And I suspect that it, or a close descendent of it, lies at the root of a great deal of the resistance to massive

modularity amongst philosophers and (to a much lesser extent) amongst cognitive scientists. Despite the fact that almost everyone now accepts the existence of unconscious mental states and processes, the picture of a diaphanous mind nevertheless maintains its grip on us, only now confined to so-called 'central cognition' or to 'personal' (as opposed to 'sub-personal') mentality.

As we saw briefly in Chapter 3.3, the human mind-reading module operates with a simplified model of the mind and its operations, included in which is the idea that the mind is transparent to itself. This is probably why the notion of unconscious perceptual states was so hard for people to accept, and met with such vigorous resistance, when proposed by Weiskrantz and colleagues (Sanders et al., 1974; Weiskrantz, 1980, 1986). And it also explains why the thesis of massive mental modularity should seem so counter-intuitive. Since the basic model employed by the mind-reading module is very likely innate (to some significant degree) it is to be expected that the intuition of a diaphanous mind would prove robust and hard to get rid of. In this respect it is similar to the intuition that it is more probable that Linda (who did voluntary work for civil rights groups and feminist organizations while at college) is now a bank teller *and* a feminist than it is that she is just a bank teller—despite the fact that it is impossible for a conjunction to be more probable than one of its conjuncts (Kahneman et al., 1982).[3]

FN:3

Those of us defending massive modularity face an uphill struggle, therefore. Just as logicians and probability theorists have to labor to get people to set aside some of their intuitions of validity and probability; and just as physicists have to work to get physics students to overcome their Aristotelian intuitions about motion (McCloskey, 1983); so massive modularists face a similar hurdle. One part of this involves convincing people that the human mind-reading faculty deploys a greatly simplified model of the mind's operations, which works perfectly well for purposes of everyday prediction and explanation, but which lacks any scientific standing (as I argued in Chapter 3.3). This has been amply demonstrated by cognitive scientists in recent decades. (See Gazzaniga, 1998, and Wilson, 2002, for recent reviews.) But note that even with conviction assured, the intuition of a diaphanous mind will be apt to reassert itself whenever we aren't paying attention. Keeping our intuitions under control

[3] Note, however, that there is a (small) element of truth in the idea of a diaphanous mind. This is that globally broadcast perceptual and imagistic states are made available to the mind-reading faculty for immediate recognition. And likewise, both mentally rehearsed actions and sentences in 'inner speech' can utilize the same global broadcasting architecture. So there is, after all, a sort of virtual central system within which experiences and thought-contents are transparently available for self-ascription and report. (See Carruthers, 2005, for discussion.) But not much of the actual work of the mind goes on within this arena.

can take constant effort. The other part of a defense of massive modularity—in addition to laying out the evidence in its support (as I have done in Chapters 1, 2, and 3)—is to show that the account can explain everything that it needs to. That is the task of the remainder of this book.

There is, of course, yet another source of resistance to massive modularity, that is so endemic that it almost deserves to be counted as part of 'common sense'. This is that the heavy dose of innateness that is part and parcel of the massive modularity thesis is inconsistent with the empiricist conception of the mind as a 'blank slate'. This picture has dominated western intellectual life since the Enlightenment, and continues to be almost a religious orthodoxy in the humanities and social sciences, and to a lesser extent in some areas of psychology. (Again Locke, 1690, serves as an early exemplar, and is generally credited as the first to introduce the metaphor of the mind as initially a blank slate, waiting to be written upon by experience.) And it is a picture that is closely intertwined with a set of 'progressive' ethical and political attitudes, to the point where denial of the orthodoxy is felt by many to be morally threatening, if not outright morally reprehensible.

The scientific case that supports an account of the mind as possessing a rich innate structure seems to me to be overwhelming, as I have argued in Chapters 1, 2, and 3. (See also Carruthers et al., 2005, 2006, and that planned for 2007.) And the remainder of this book is designed to reply to the main intellectual challenges to such an account. But the *moral* objections will have to pass unanswered. Replying to them would require quite a different sort of book. And that book has, in any case, already been written—see Pinker (2002), who does a masterful job of identifying the various political and emotional currents underlying blank-slate psychology, while at the same time disarming them of intellectual credibility.

1.2 Massive Modularity and Flexibility

Should massive modularity theorists predict that there will be limitations on the degree of *flexibility* of the human mind, and its resulting behavior? A number of different kinds of flexibility and inflexibility can, and should, be distinguished. One kind of flexibility is flexibility of *action*. As we saw in Chapter 2.7, the minds of many animals are inflexible in the following sense: particular types of desire can only be paired with certain kinds of belief, and not others; and particular types of desire can only recruit certain kinds of action schema, and not others. But as we saw in Chapter 2.8, this sort of inflexibility can be overcome by minds that have a capacity for the mental rehearsal and subsequent global broadcasting of action schemata. The human mind undoubtedly has

such a capacity, as do probably the minds of other species of ape and earlier species of hominid.[4] But such a capacity doesn't by any means compromise the massively modular status of the mind; on the contrary, it recruits the activity of existing modules to subserve the task of action selection.

All this will loom large again in Chapter 5, when we discuss the distinctive *creativity* of the human mind. But it is worth noting at this stage that there are independent reasons to think that feedback loops of various sorts are the right place to look for sources of creativity and flexibility. (I owe this observation to Chris Pyne.) For consider what happens in video feedback (Crutchfield, 1984). If you direct a video camera at a blank television screen, in circumstances where the camera is wired up so that its output will be displayed on that very screen, then all sorts of interesting things start to happen. Rich patterns of color and shape tend to result from the continual cycling of the feedback loop alone, without the injection of any initial content, and without design. So, too, perhaps in the human mind, once it begins its cycles of mental rehearsal.

Another kind of flexibility is that a mind (and the behavior in which it results) might be more or less *context-sensitive*; and yet another is that it might be more or less *stimulus-bound*. (Other forms of flexibility will be distinguished shortly.) The minds of insects seem inflexible in both of these sorts of ways. A wasp might continue with the same kind of behavior (building a nest to protect her eggs, say) in the same kind of situation (pregnancy) irrespective of the context in which the activity is taking place, such as the presence of a human experimenter who makes holes in the mud tube of the nest, or who buries the nest in sand. (See Gould and Gould, 1994, on the Australian digger wasp.) Likewise an insect might always behave in the same way when presented with the same stimulus, irrespective of circumstances.

It is obvious that human minds aren't inflexible in either of the above senses, however. One of our most distinctive properties is the way in which we can adapt (not always, but at least sometimes) to changed circumstances, and think and behave in a context-sensitive manner. And likewise our thought processes plainly aren't inflexible in the stimulus-bound sense, either. On the contrary, we routinely entertain thoughts, and whole sequences of thought, that bear no relation to our current physical or social circumstances. I shall discuss in turn the alleged challenges to a thesis of massive modularity raised by these forms of flexibility.

[4] Geary (2005), too, emphasizes the importance of mental rehearsal in explaining distinctively human problem-solving abilities. But he thinks that a capacity for mental rehearsal is restricted to humans. I think, on the contrary, that it is present in other apes, but greatly enhanced in humans. See Chapter 5 for further discussion.

1.3 Context-Flexibility

I don't believe that the context-sensitive form of flexibility raises any particular problem for a massively modular conception of the human mind. For there are a number of distinct but mutually compatible strategies that a massive modularist can adopt in seeking to explain our distinctive context-flexibility. In order for us to see this clearly, however, it is necessary to distinguish between two different *forms* of context-flexibility. One way for an organism to be context-flexible is for it to pick up on the information in the environment that is relevant to its current goals, and to modify its behavior accordingly. This would lead us to expect that different organisms with the same goals in the same circumstances should behave similarly. But another way in which organisms can be context-flexible is where different individuals are apt to pick up on and respond to different aspects of the context, leading those individuals to behave *differently* in the *same* circumstances. (Human beings are context-flexible in both of these senses, of course.)

When context-flexibility is construed in the first of the above ways, it should be obvious that it raises no special problems for massively modular conceptions of mind. Quite the contrary. For the greater the number of modules that exist, and that are operating in parallel, the more features of the environment / context the agent can pick up upon and respond to. And so we can, in effect, turn the objection on its head. A monolithic mind containing just one general-purpose processing and inferential system (if such a thing can really be envisaged) would surely be a mind that could only pick up on one item of information at a time, or that would at least be limited in the flexibility that it displayed in relation to features of context.

If there is a problem for massive modularity arising out of the context-sensitivity of the human mind, then, this must be a problem in respect of context-flexibility of the second sort, pertaining especially to differences *between* individuals. And it is easy to see how the objection might go. For shouldn't we expect the outcome of the operations of the same set of modules (especially if innate) to be the same whenever presented with the same input? There are, however, at least three different, but mutually consistent, sorts of response that a massive modularity theorist can make.

1.4 Three Ways to be Context Sensitive

One way of reconciling massive modularity with the context-sensitive character of the human mind (in both of the above senses) is proposed by Sperber (2005). He argues convincingly that the operations of the mind as a whole should be characterized by various kinds of *competition* amongst modules. Modules will compete with one another for a variety of forms of resource, both physical

(such as increased blood-flow to one region of the brain rather than another) and cognitive. Amongst the latter might be included competition for various forms of working memory, and for a variety of kinds of attention. Moreover, where a module can receive input from a number of other systems (in the way that we described for the practical-reasoning modules in Chapter 2.8), then there might be competition amongst those systems to have their outputs received and processed as input by the module in question.

On this approach the context-sensitivity of a massively modular mind might be expected to arise in something like the following manner, then. Different modules are cued by different features of the environment—social, physical, animal, vegetable, etc.—and at various levels of abstractness (e.g. suddenly moving stimuli and loud noises versus cheater detection). All, when activated, compete with one another for resources, and to get their outputs entry into downstream inferential and decision-making systems. But how this competition pans out in any given case might often be highly sensitive to the details of the context (both environmental and cognitive), and also to the learning history of the person in question. Certainly there should be no hint of any crude environmental determinism here.

A second proposal for dealing with the context-sensitivity issue comes from Barrett (2005). He elaborates and discusses what he calls 'the enzyme account' of modularity. The idea is to model the operations of modules on the way in which enzymes build proteins within cells. There are many different kinds of enzyme within a cell. Each has a characteristic shape, and floats around waiting to meet a protein that matches that shape. When it finds one, it builds a new protein of a characteristic sort and pushes the result back out into the soup of chemicals within the cell once again. Translated into cognitive terms, the idea is that there might be a whole host of specialist processing devices ('modules') all focused on a common 'bulletin board' of representations. Whenever a device comes across a representation that 'fits' its input condition it gets turned on, and it then performs some set of transformations on that representation before placing the results back on the bulletin board for other devices to pick up upon.

One can thus envisage a cascade of inferences and transformations taking place (with some modules looking for representations that possess increasingly abstract tags placed there by other modules, and so forth), but without there being any architectural constraints on the flow of information through the system. And the result would be processing that is highly context-sensitive, but resulting from the independent operations of a set of enzyme-like modules, whose collective output depends partly on happenstance.

This proposal fits nicely with the 'global broadcasting' model of perceptual consciousness put forward by Baars (1988, 1997), in support of which there is

robust empirical evidence (Baars, 2002, 2003; Dehaene and Naccache, 2001; Dehaene et al., 2001, 2003; Baars et al., 2003; Kreiman et al., 2003).[5] We can think of the enzyme model as an account of how the various conceptual modules continually scan the contents of globally broadcast states, searching for ones that trigger their input conditions. Indeed, we can think of it as a model of how the conceptualization of perceptual states takes place, given that the concepts in question are deployed by specialist modules of one sort or another.

The enzyme model looks plausible as an account of how perception gets conceptualized by modular processes, then. But it might seem singularly *im*plausible as an account of how more abstract modules operate, such as the social contracts / cheater detection system (Cosmides and Tooby, 1992). For *cheat* isn't a perceptual category. In which case, for the account to work, it might appear that we would need to postulate multiple global broadcasting systems, some dealing with perceptual contents, and some with more abstract conceptual ones. Yet there is no independent evidence that the latter exist.

I doubt that this is a serious problem for the enzyme model, however. Granted, *cheat* isn't a perceptual category. But why shouldn't that concept nevertheless become attached to a perceptually represented and globally broadcast item? In effect, what would be globally broadcast would be a perceptual item conjoined with the thought, *that is a cheat*, where the indexical 'that' refers to the perceived person in question. This combination could be made available to a wide range of consumer systems (enzyme-like modules), some of which might be searching for the content *cheat* in order for their processing to be turned on.[6]

The enzyme model doesn't just provide us with an account of the conceptualization of perception, however. It also suggests how module-generated predictions based on current perception can also be made globally accessible, utilizing the back-projecting neural pathways present in all perceptual systems to create visual and other forms of imagery, which can then be globally broadcast in turn. Suppose, for example, that I see a ball flying towards a glass window, from which my physics module predicts that the window will shatter.

[5] In saying that there exists robust evidence of global broadcasting, I don't mean to say that there is evidence supporting global broadcasting as a proposed reductive account of phenomenal consciousness. For it may well be the case that mammalian brains all share a global broadcasting architecture in respect of a privileged set of perceptual states; and it may well be the case that those states are in fact phenomenally conscious in humans; but they might not be phenomenally conscious *because* they are globally broadcast; indeed, on my own account, they aren't (Carruthers, 2000, 2005).

[6] Such a model provides us with a natural way to think of the operations of the natural-language comprehension system, indeed. That system operates on perceptual input to deliver a conceptual output of the message being communicated. But the latter isn't detached and independent of the initial percepts. On the contrary, the phenomenology of speech perception is that one *hears* the meaning of the words being uttered, as well as hearing the sounds—pitch and tone of voice and so forth—that constitute the utterance.

This prediction can then be displayed in the form of a visual image prior to the event occurring. This makes that prediction widely accessible to the full range of central / conceptual systems, many of which wouldn't normally have received the physics module's output directly. These systems can then generate yet further inferences or emotional reactions that might prepare me for action (running from the scene, perhaps, if I was the hitter).

At a much more basic level than either the competition-for-resources or the enzyme-model responses to the problem of context-flexibility, however, it should be emphasized that many modules are *learning* systems, and that many other modules are in the business of *building* modular systems from the contingencies of environmental interactions. So overall flexibility of behavior—both in response to variations in the natural and social environment, and co-varying with the different learning histories of different individuals—is precisely what a massive modularist should predict. (And in addition, of course, there will be innate differences between different individuals concerning the properties of their respective learning modules, yielding yet further differences in behavior.) Let me elaborate.

As we emphasized in Chapters 2 and 3, many of the modular systems that constitute the minds of both animals and humans are designed to extract information of some specific sort from the environment. The multiple systems involved in spatial navigation are designed to extract information about the spatial relationships between the agent and other things, and amongst those things themselves, for example. (Likewise the human language faculty is designed to extract information about the meanings of utterances spoken by members of the agent's local community.) How a creature will navigate will then be sensitive to the spatial context in which it has done its learning, and in which it now finds itself. (Likewise the language that a person speaks will be sensitive to the linguistic context in which that person has been immersed.)

Not only do humans have what Fessler (2006) calls 'information-rich' learning modules of the above sort, but they also have information-*poor* learning systems, as we saw in Chapter 3. A number of investigators have demonstrated that humans have a variety of dispositions that aid in the learning of culture-specific information, where what is to be learned can't be second-guessed by the evolutionary process (Richerson and Boyd, 2005). People have a disposition to observe and to copy those who are prestigious, together with an associated emotion system that generates admiration (Henrich and Gil-White, 2001). And they have a disposition to observe and copy slightly older / more experienced peers who are similar to themselves along some relevant dimension. Moreover, they have a disposition to observe, learn, and attach intrinsic motivation to the norms that are current in their community (Sripada and Stich, 2006).

These dispositions, together with the background capacities underlying imitation with which they interact, make possible the development of rich technological and normative cultures. And again, a massive modularist should predict that the configuration of any given individual's mind will be sensitive to the context of the surrounding culture, with wide variations in outcome (even within a single culture) depending upon the happenstance of details of the individual's learning history, and on variations in learning strategy.

In addition to acquiring *knowledge* from the surrounding culture, of course, humans also acquire a range of behavioral skills, from stone-tool knapping, through cooking, to kayak building, to reading and writing. In each of these cases it is plausible to claim that what are being assembled during the learning process are behavioral modules, which can be held constant when yet other behavioral modules are acquired, and whose properties can vary, and can be influenced, independently of the others. Into the process of assembly will go observation, practice, sometimes explicit instruction, and feedback of various kinds. And again the result will be a behavioral repertoire that is highly context-sensitive, and that will vary with the individual's physical and cultural environment, as well as with the details of their idiosyncratic learning history.

In summary of this sub-section, then, I believe that massive modularity theorists have a number of resources with which to explain the distinctive context-sensitivity of human cognitive processes and behavior. Context-sensitivity doesn't present an especially difficult challenge for the thesis of massive modularity to meet.

1.5 The Stimulus-Free Mind

In contrast to context-flexibility, the stimulus-free nature of much human mental activity does pose more of a problem for massively modular conceptions of mind (as it does for anti-modular accounts as well; the problem is by no means unique to modularity theory). But here, too, we can distinguish between two different forms of stimulus independence. One of these is relatively straightforward to explain, and will be tackled in Section 2. The other is much harder, and will be deferred to the closing sections of Chapter 5, when we complete our discussion of creativity.

One challenge is to explain how a network of belief-generating modules and desire-generating modules can be arranged into an architecture in such a way that the behavior of the whole system can often be free of environmental input. Somehow we will have to provide for the overall system to be *self-stimulating*, or at least *self-sustaining*, in its operations.

Recall how human thought-processes can be radically independent of current circumstances. I can get into a day-dream and spend minutes or hours

reliving events from my past, or fantasizing about my next vacation. Or I can be sitting immobile at my desk thinking about what I should say during an annual appraisal interview with my boss some weeks in the future. These obvious facts might appear to present something of a problem, because the basic model sketched in Chapters 2 and 3 is a *feed-forward* one (see Figure 4.2). External stimuli are processed by the perceptual systems, and the resulting percepts are made available to a range of belief-forming and desire-forming modules; the ensuing mental states are made available to practical reasoning, which issues in an act or in an intention to act. It is initially hard to see what scope there can be for these systems to operate in the absence of, or independently of, any sort of perceptual stimulus.

Many neural systems contain back-projecting neural pathways of various sorts, of course. This is certainly true of the visual system, where there a pathways projecting all the way back to the primary cortical projection area V1. These are used to direct attention and to 'query' degraded or ambiguous input; and they are also the basis of visual imagery, as we saw in Chapter 2.2 (Kosslyn, 1994). Moreover, humans (and other apes) have a capacity for mental rehearsal of action schemata, which takes a representation near the 'output' end of the mind (in motor control) and uses it to build a quasi-perceptual representation of the intended action, which can then be globally broadcast and received as input by the full suite of central / conceptual modules, as we saw in Chapter 2.8. Showing how these elements can be utilized and combined to give rise to the distinctive stimulus-independence of human thought will be one of the tasks of the present chapter, to be undertaken in later sections.

There is another way of characterizing the stimulus-free character of the human mind which is much more deeply challenging, however. This is the property that was at issue in the famous debate between Chomsky (1959) and Skinner, which Chomsky (1975) has since taken to describing as 'the creative aspect of language use', or 'CALU'. Confronted with one and the same external stimulus (a painting hanging on a wall), there are no end of things that one could intelligibly say. One might say, 'Dutch', or, 'It is hanging too low', or, 'It clashes with the wallpaper', and so on, and so on, without limit. Each of these responses might be perfectly *appropriate* in the context, but without being under stimulus control. How is this possible?

Some aspects of this problem reduce to the problem of explaining the context-sensitive character of human thought and behavior (discussed in Section 1.4), and can be handled accordingly. Thus it is certainly to be expected that different people, with their different and idiosyncratic learning histories, might respond differently to one and the same stimulus. And even for the same person at different times, one might expect that the competition

between modules to get their outputs entry into the language production system might pan out differently, depending upon different motivational and contextual saliencies. But it is very doubtful that the creative aspect of language use can be exhaustively explained in either of these ways.

It should be stressed, however, that the creative aspect of language use isn't just a problem for massive modularists. On the contrary, it is a problem for everyone. And Chomsky (1975) has even suggested that the problem may be so hard that its solution is *cognitively closed* to us, in the same sort of way that an explanation of gravitational phenomena is cognitively closed to a rat. It will therefore be a large 'plus mark' in favor of massively modular approaches to cognition if they can enable us to make some progress with the problem. I shall return to this topic towards the end of Chapter 5.

1.6 Flexibility of Content

Should massive modularity theorists expect that there will be limitations on the flexibility with which concepts (the components of thought contents) can be combined? (Call this 'content-inflexibility'.) I believe that the answer to this question is 'Yes'. There are two reasons for this. The first is that we surely shouldn't expect that every system will be connected up with every other. (See Figure 1.3 and the surrounding discussion in Chapter 1.) The flowchart of information through the mind to the point of decision-making should place some restrictions on which concepts can be combined with which, and when. So if one concept can be proprietary to one conceptual module and another to another, then these might be two concepts that can never get combined into a single thought. This will be because the modules that initially generate tokens of those concepts lack any connection with one another, direct or indirect.

This problem would certainly be mitigated if there were some sort of domain-general formal logic module, as we speculated in Chapter 1.2 that there might be. For this would be capable of taking any (small) set of beliefs produced by any given subset of modules and deducing some of the simpler logical consequences from those beliefs. So it ought, in particular, to be capable of taking any belief *P* and any belief *Q*, and combining them to form the cross-modular thought, *P and Q*. For this requires only a simple step of conjunction-introduction. Likewise if there should turn out to be a module capable of calculating the statistical dependencies amongst arbitrary pairs of properties, then it would be capable of generating a proposition of the form, $P \supset Q$, for any *P* and any *Q*.

Notice, however, that neither of these proposals would make it possible for two module-specific concepts to be combined within a single *atomic* (as opposed to molecular or quantified) proposition. For example, if the output

of some sort of geometric module were the belief that a target object is in a corner with a long wall on the left and a short wall on the right, and the output of some kind of object-property module were that the target is near a red wall, then there might still be no way for these two beliefs to be combined into the single integrated representation, THE OBJECT IS IN A CORNER WITH A LONG WALL ON THE LEFT AND A SHORT *red* WALL ON THE RIGHT, even if the short wall *is* a red wall. The best that the logic-module would be able to get us is the thought, THE OBJECT IS IN A CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE RIGHT *and* THE OBJECT IS NEAR A RED WALL.

Without knowing the details of the flowchart of modular connectivity in the mind, however, it is hard to generate specific predictions from the claim that a massively modular mind should display content-inflexibility, except in obvious cases. Thus we can predict that contents concerning surface boundaries produced early in the visual system, for example, shouldn't be combinable with thoughts about the stars, nor about other people's beliefs.[7] But as for which concepts might fail to be combinable with which other concepts, this is impossible to predict without knowing the connectivity of the conceptual belief-generating modules in question.

FN:7

The second reason for expecting that a massively modular mind should be to some degree content-inflexible is as follows. Even if two concepts can be combined somewhere for one purpose, it doesn't follow that they can be so combined for another, or that the system that is operative in the latter context can access the combined representation. So for concreteness (and in line with the evidence briefly reviewed in Chapter 2.3), suppose that the representation produced by the geometric module, THE OBJECT IS IN A CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE RIGHT, and the representation produced by the object-property module, THE OBJECT IS NEAR A RED WALL, are routinely passed to a map-creating system, which builds the integrated representation, THE OBJECT IS IN A CORNER WITH A LONG WALL ON THE LEFT AND A SHORT *red* WALL ON THE RIGHT. But under conditions of disorientation (when the goal is to find out where I am when I am lost), the latter representation isn't accessed. Rather, the reorientation goal mandates a search for geometric information alone, pulling up the geometric representation, THE OBJECT IS IN A CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE RIGHT.

[7] Of course one can, downstream of the visual system, come to conceptualize something *as* a boundary of a surface, and then go on to wonder whether the stars have boundaries like that. But this is another matter. Here one re-represents, in fully conceptual format, a content similar to one that had elsewhere been deployed within the visual system. This isn't the same as saying that the latter content has been extracted and combined with a thought about the stars.

In such circumstances there might be no way for the content-integrated representation, containing *red*, to have any effect on behavior. So the organism would display a sort of content-inflexibility here, but only in conditions of disorientation, not absolutely. In principle it might be possible to solve this problem via mental rehearsal. For example, if one were mentally to rehearse going towards the red wall, then the geometric system could perhaps kick in with the actual location of the object. But in any given case such rehearsals might lie outside the normal repertoire of agents, with their acquired strategies and heuristics for generating useful rehearsals. It might just never occur to people to rehearse turning towards red, for example.

In this sort of case, too, however, it is hard to generate specific predictions without knowing the identities of the modules that make their outputs available to the practical-reasoning systems, and the manner in which the latter operate. But we do at least have some reason to think that the example of local content-inflexibility just sketched is real enough, since it is displayed in the behavior of rats and young children. This evidence will be extensively discussed in Section 4.

Even if the prediction of content-inflexibility is vague and unspecific, however, it still creates a significant problem for the thesis of massive mental modularity. For, in contrast to that prediction, humans would appear to be capable of freely combining concepts across the boundaries of all central / conceptual modules. This is manifest to ordinary introspection. I can be thinking about thoughts one moment, horses the next, and then a landslide the next; and I can then wonder what led me to think about thoughts, horses, and falling stones—thereby combining into a single thought concepts drawn from the domains of folk-psychology, folk-biology, and folk-physics. And likewise for any set of conceptual modules that you care to mention. How is this possible, unless there is some a-modular central arena in which the contents of conceptual modules can be received and recombined, further inferences drawn from the results, and so forth? This is yet another challenge to massive modularity.

It might be replied that I have, over the previous two chapters, committed myself to the existence of just such a central arena, namely the practical-reasoning system (see Figure 4.2). I have suggested that this system is capable of receiving any desire and any belief as input. So why shouldn't it be this system that has the power to combine and recombine concepts drawn from disparate domains? For it can at least *receive* all of those concepts amongst its inputs. But in the first place, practical reason actually consists of many different desire-specific modules, as we saw in Chapter 2.7. And even though mental rehearsal and self-monitoring can transform the collective operation of these modules into a more-unified overarching practical-reasoning system (as we saw

in Chapter 2.8), this couldn't help with the first problem of content integration outlined above (where two modules are isolated from one another absolutely); and it might not be sufficient to solve the second, either, as we have just seen.

Moreover, task-analysis of the requirements of practical reason (at least of the sort found in non-human animals) suggests that the combinatorial and inferential powers of the practical-reason system should be quite severely restricted.[8] While practical reason can receive any desire and any belief as input, it should have no capacity to conjoin and integrate the contents of such states, except where the beliefs in question are conditional in form—in which case it might have the power to collapse P ⊃ Q and Q ⊃ R to form the conditional P ⊃ R, thus 'conjoining' the propositions P and R together in a single thought for the first time. Nor should it have the capacity to draw many inferences from the propositions it receives, except in so far as it executes the practical reasoning equivalent of *modus ponens* (namely: *I want R, P ⊃ R, P is something that I can do, so I'll do P*).

It might be claimed, of course, that precisely what happened in the evolutionary transition from our great-ape ancestors to ourselves was that the practical-reasoning system underwent a transformation into a general content-conjoiner and inference engine. But such a proposal remains mysterious in the absence of (a) a more detailed task-analysis of the functions that such a transformed practical-reasoning faculty might be expected to perform, and of (b) some account of the evolutionary pressures that would have led such changes to occur. And as it stands, the proposal seems inconsistent with the complex and hedged-about 'one function / one module' generalization that emerged from our discussions in Chapter 1.

1.7 Flexibility of Reasoning Process

I have argued that a massively modular model of the human mind might lead us to predict (falsely, of course—hence our problem) that there are some constraints on the mind's capacity to conjoin and combine concepts drawn from different modular systems. Let me now argue that the same model should lead us to expect that there might be severe limits on the kinds of reasoning

[8] And even though human practical reasoning is by no means restricted in the manner in which thought-contents can be combined and conjoined, this doesn't solve our problem. For generating intentions and actions from beliefs and desires is one task, combining and drawing other sorts of inferences from thought-contents is quite another. So the argument from design, articulated in Chapter 1, should lead us to predict that there should be distinct systems for these distinct tasks. At the very least we are owed an account of how a simple practical-reasoning system could evolve into some sort of universal content-conjoiner. The topic of human practical reasoning will be further pursued in Chapter 7.

and decision-making *process* in which humans can engage. This will give rise to yet another challenge for a thesis of massive modularity to overcome. But first I need to provide a little background, in order that the argument can be set up.

If we accept that there is one overarching decision-making / practical-reasoning system (albeit made up out of multiple sub-modules), then this will be a point at which 'everything comes together.' The decision-making system is the point of maximum convergence of information, since the outputs of the various belief-modules and desire-modules should all be made available to this system. (Indeed, all may be competing to have their outputs received as inputs to it; see Section 1.1 of this chapter.) There will therefore be maximum demands on the computational resources of the practical-reasoning system. If its computations are to be tractable, and executable in real time (sometimes in fractions of a second), we might expect it to deploy a number of heuristics, or 'quick and dirty' short-cuts, in order to ease the computational load, and to render practical reason's task more frugal in terms of time and resources.[9]

FN:9

As it turns out, this is an idea that has been explored with remarkable success by Gigerenzer and colleagues (1999). They propose and examine a range of decision-making heuristics of a simple sort, pitting them against a number of much more sophisticated competitors such as multiple regression and Bayes' rule. It turns out that under a variety of test conditions (both real-world and simulated) such simple heuristics are almost as reliable as their fancier competitors, and are much more frugal in terms of time and computational resources. However, there are also predictable kinds of circumstance in which such heuristics will go wrong (Kahneman et al., 1982).

I postulate that three different *kinds* of heuristic should be employed. First, there should be heuristics governing *how long* one should search for information and reason about the alternatives before taking a decision. The mate-choice heuristics explored by Gigerenzer et al. (1999) would be an example. Secondly, there should be heuristics governing what sorts of information one should seek out and rely on in a given choice situation. For example, a variety of data suggest that when animals are disoriented they employ a set of nested heuristics for deciding in which direction to travel to reach their target. And for many species of animal, including rats and human children, those heuristics are organized in the following sequence: (1) seek a directional beacon (e.g. the

[9] Notice that we certainly should *not* expect practical reason to be performing calculations of maximum expected utility, integrating measures of the degree of desirability of all goals with measures of the likelihood of all foreseeable ways of achieving them. Although philosophers and economists routinely assume that maximizing expected utility is a normative constraint on human practical reasoning, this is plainly a mistake, provided that one accepts the traditional principle that '*ought* implies *can*.' See Gigerenzer et al., 1999.

sun, or a distant landmark such as a familiar line of hills); (2) if no beacon is available, look to the geometrical properties of the local environment and seek a match with geometric memory; (3) if no geometric match is found, seek a recognizable local landmark and attempt to locate its position on a mental map (Shusterman and Spelke, 2005).[10]

FN:10

The second of the above heuristics concerns the type of information that one should search for in order to reach a given decision; the first concerns how long one should search before reaching a decision (or abandoning the task). To see a place for a third type of heuristic, recall that in apes the practical-reasoning system will also have a capacity for mental rehearsal of action, feeding a sensory (normally visual) representation of the action-to-be-considered back through the various belief- and desire-generating modules as input, and monitoring one's bodily / emotional reactions to the results. The obvious type of heuristic to expect here would concern which of the action-schemata in one's action database to activate in a mental rehearsal. Heuristics such as 'Take the Last' (i.e. activate the action-representation that was used last in connection with a decision-problem of this type) explored by Gigerenzer et al. (1999) could naturally be adapted to serve in this third kind of role.

While it is now well-established that humans do use a variety of decision-making heuristics (Kahneman et al., 1982; Gigerenzer et al., 1999), just as a massive modularity thesis might predict, it is equally obvious that human beings aren't limited to, nor strongly constrained by, those heuristics. Courses in logic, or in probability theory, or in scientific method really can make a difference to the ways in which people think and reason, at least when they reason reflectively. By acquiring beliefs about the ways in which we *should* reason, it is possible for us to change the ways in which we *do* reason, at least some of the time. This gives rise to yet another challenge for a thesis of massive mental modularity to answer. How can the operations of a range of inferential modules be overridden by our explicit beliefs about norms of reasoning? By what mechanism can the latter pre-empt or control the former?

1.8 More Challenges: Creativity, Science, and Practical Reason

We have discovered, then, three different forms of flexibility that look problematic from the perspective of a massively modular conception of the human mind: stimulus-independence, flexibility of content, and flexibility of reasoning and decision-making processes. None of these challenges presents a problem of *principle* for massive modularity, of course, of the sort that Fodor (2000) attempts

[10] In other species of animal, including monkeys and chickens, the ordering of (2) and (3) appears to be reversed (Vallortigara et al., 1990; Gouteux et al., 2001). I shall return to this point in Section 4.

to defend. Rather, they are just that: *challenges*. We are challenged to explain how humans manage to attain the flexibility of thought and reasoning that they manifestly possess, supposing that the thesis of massive modularity is true. Answering these flexibility-challenges will form the topic of the remaining sections of present chapter.

In addition to the flexibility-challenges to massively modular accounts of the human mind, of course, there will remain the problem of explaining the distinctive *creativity* of human thought processes. This is manifested in childhood pretend play, in story telling and fantasy, in metaphor, and in science. There is a sense in which creativity might perhaps be thought of as a sub-species of content flexibility, since in all these different domains creativity can manifest itself in the formulation of novel thoughts, not plausibly produced by the routine (module-dependent) processing of perceptual input. However, creative content-generation raises problems of its own, as we shall see. Accordingly, it will be given separate treatment in Chapter 5.

While scientific thought and reasoning may depend in part upon our creative abilities, they involve much else besides. Indeed, they provide a significant challenge for *any* account of the mind to explain, whether modularist or anti-modularist. Although disagreeing about almost everything else, for example, Pinker (1997) and Fodor (2000) are united in thinking that the capacity for scientific reasoning is a genuinely *hard* problem for any scientific account of the mind to explain, to be likened to the so-called 'hard problem' of consciousness (Chalmers, 1996). This is a challenge that I shall return to in Chapter 6. It will be an important point in favor of modular models of the human mind if they can demonstrate progress towards meeting it.

Finally, there remains the problem of explaining distinctively human practical reasoning. There seem to be no limitations on the kinds of consideration that can enter into such reasoning, and new strategies for practical reasoning can be learned. Once again there is a sense in which this can be seen as a sub-species of the various flexibility-challenges to massive modularity that are under consideration in the present chapter; and both creativity and (I shall argue) science-like reasoning are presupposed. But here, too, it will turn out that there are distinctive problems to be addressed by a massively modular account. Further discussion of this topic will therefore be deferred to Chapter 7.

2 Stimulus Independence and Inner Speech

Recall from Chapter 2.8 that one of the precursor systems within pre-hominid forms of cognition was a capacity for mental rehearsal of action. This sub-divides

into two parts: a capacity for creative generation of action schemata when problem solving; and a capacity to map those action schemata onto an appropriate perceptual representation of the action, so that a representation of the action being contemplated can be received as input by the various central / conceptual modules in such a way that its consequences (both physical and social) can be calculated. I shall first discuss how such a capacity may be deployed in visual and other forms of imagery, before turning to its more-novel manifestation in 'inner speech'. Thereafter I shall devote some time to discussing a third category of stimulus-independent cognition (one that is independent of *current* stimuli, at any rate), namely episodic memory.

2.1 Imagined Actions and Sequences of Imagery

Recall the example of Belle from Chapter 2.8, who was faced with a problem. Each time she went to retrieve some hidden food whose location she had been shown, the alpha male would follow her and push her aside as soon as she began digging, seizing the food for himself. And recall, too, how she may have been able to hit upon her solution. Mentally rehearsing her initial intention (to go to the food), she imagines failure, and feels disappointment. So she activates and mentally rehearses some of the other actions available to her, from one of which (digging elsewhere) she is able to envisage circumstances (the male's preoccupation) in which she can imagine obtaining and eating the food. This gives her a novel two-step plan which she implements successfully.

This is a sequence of cognitive activity that is prompted by an external stimulus, perhaps (e.g. the opening of the enclosure in which the chimpanzee knows the location of some food), but which thereafter proceeds in large degree independently of external stimuli. And in humans, with their much-increased conceptual resources and much-increased capacity for reasoning and prediction, we might expect that such sequences of planning by mental rehearsal should become considerably more frequent and extensive. And indeed, human subjects can spend extended periods of time physically inactive, trying out in imagination a variety of scenarios for solving some problem or for achieving some goal. This is one aspect, at least, of the stimulus-independence of human cognitive processes.

In order to go further in the direction of stimulus-independence, we need only suppose that humans have acquired the disposition to activate action schemata that are much more loosely connected to, and/or less controlled by, external stimuli. (As we shall see in Chapter 5, such a disposition may be an effect—and perhaps the proper function—of childhood pretend play.) The shape of a passing cloud, a phrase overheard on the bus, or a note in a diary can all prompt one to engage in extended periods of action-schemata rehearsal,

sometimes in the service of medium- or long-term planning, sometimes serving as mere fantasy. And all of this can take place wordlessly, in sequences of visual and other images generated by feedback loops from activated action schemata. Each image in the series is taken as input by the various central / conceptual modules, which generate further predictions and emotional reactions. And this can serve as an internally generated stimulus for the activation of yet another action schema, which gives rise to yet another sequence of images, and so forth.

2.2 Inner Speech

Speech, of course, is a form of action. So one would expect that the precursor capacity for action rehearsal would carry over into this new domain. In which case there should be a capacity for creative generation of speech schemata (i.e. representations of possible utterances in a code appropriate for receipt by the motor systems), and a capacity to map those representations into a sensory modality (normally hearing), so that they can be globally broadcast and received as input by the conceptual modules. Indeed, we can predict, I think, that there would have been *special* pressure for the development of such feedback / rehearsal loops in the course of the evolution of language. Let me elaborate.

There are two main accounts of the original functions of language, and of the evolutionary pressures that led to its development. (Another will be mentioned later.) One is that language was for mutually beneficial exchange of adaptive information, in which case its evolution is to be explained in the same general way as the evolution of reciprocal altruism (Pinker, 1994; Sober and Wilson, 1998). Language was (and is) a way of transferring information from one person to another that is almost cost-free for the donor, but which can bring huge benefits to the receiver. If I have seen a poisonous snake wriggling into your hut, then telling you about it costs me only a few moments of my time and a few extra breaths; whereas it may save you your life.[11]

FN:11

The other main proposal is that language evolved initially for social functions. Language was for *gossip* (Dunbar, 1996), which served sometimes to maintain and strengthen alliances and personal relationships, sometimes to manipulate other people; as well as functioning as a powerful mechanism of social control. According to Dunbar, gossip is a way of *grooming* other people, enabling humans to maintain personal relationships in larger groups than those existing in any other primate species. But it would also, and very rapidly, have become

[11] If you are my rival, of course, then telling you about the snake *will* have a significant cost—it will be the cost of foregoing an opportunity to get rid of a rival. The example shows that the costs and benefits of exchanging information will be by no means always easy to calculate.

a mechanism for achieving many other social ends, from wooing a mate to deceiving a rival (Miller, 2000). And it is also the primary means of enforcing social norms (Sripada, 2005). If someone has broken a social norm, then gossiping about it can lower their social standing, leading in extreme cases to social exclusion.

It may well be the case that *each* of these accounts is correct, and that language served *both* informative *and* social functions from the very beginning. But even if language started out in just one of these ways, it would rapidly have become co-opted for the other. As soon as you can inform people of things then you can start using language both to maintain alliances and to deceive and manipulate people, provided you are socially smart (as our hominid ancestors no doubt were). And as soon as you have a language that is rich enough to gossip about a variety of kinds of social activity, then you will have a language that is rich enough to exchange other forms of information as well.

The important point for my purposes, however, is this. Whichever of the above accounts is correct, there would have been intense pressure for the development of mechanisms of mental rehearsal of speech, leading to the sort of architecture depicted in Figure 4.3. Rehearsal provides a plausible mechanism for how one can come to predict the likely consequences of saying, 'There is a black mamba in your hut', for example, thereby discerning that this is an utterance worth making. For this would enable the mind-reading system to receive as input a representation of that utterance being made, and to generate the prediction that you will either stay out of your hut or enter it armed with a stick. And likewise mental rehearsal of an utterance would

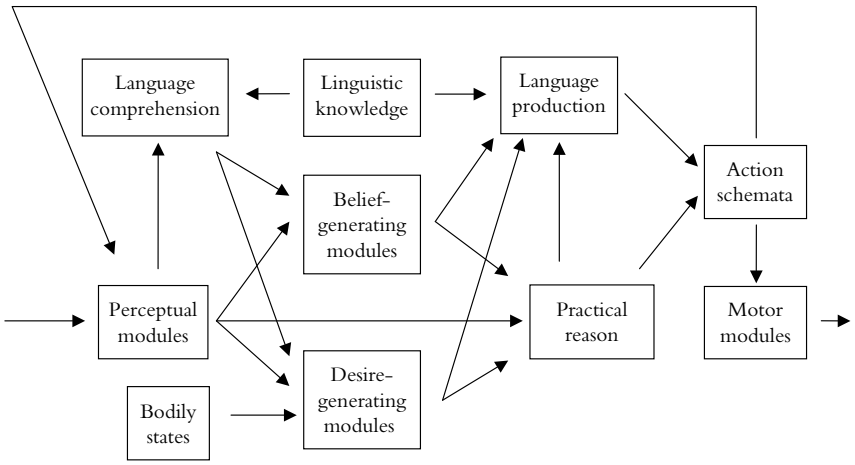

Figure 4.3. Mental rehearsal of speech

explain how one can come to predict the consequences of saying to a potential mate, 'Your eyes are like sapphires'—again discerning that this might be an utterance worth making—thus enabling the mind-reading system to generate the prediction that it would cause her to be pleased, and to be more receptive of one's advances.

Note that in each of these cases it is the mind-reading system that is one of the crucial consumer-systems for the mentally rehearsed utterance. Gomez (1998) makes a very plausible case that there would have been a sort of evolutionary 'arms race' in the development of both language and mind-reading, with advances in either one putting extra pressure on the development of the other. Moreover, Sperber and Wilson (2002) argue that there is a distinct sub-module of the mind-reading faculty devoted to communication, and to the calculations of *relevance* that underpin successful communication. (Consistent with this idea, Happé and Loth, 2002, show that children make allowances for the false belief of another in communicative contexts before they are capable of solving false-belief tasks of the regular sort.) So it may be that in modern humans there is a distinctive form of speech rehearsal that utilizes only a feedback loop to this specialized system, which one can deploy swiftly and unconsciously in the course of much regular communication, only resorting to full-blown conscious rehearsal where wider cognitive effects need to be predicted.

With a language faculty in place, together with a sophisticated capacity for mental rehearsal of linguistic utterances, then all that would have been needed was the evolution of a disposition to generate utterances *outside of* any communicative context. All that had to happen, in effect, was to take an existing network and use it more often. We would then get the cycles of 'inner speech' that are such a characteristic feature of human waking life (Hurlburt, 1990, 1993).[12] Such a cycle will begin with a mentally rehearsed utterance, perhaps primed by something recently seen or heard. That utterance is then globally broadcast by the auditory system and received as input by the language comprehension system. The latter generates from it a propositional representation—perhaps building

FN:12

[12] Subjects in Hurburt's studies wore headphones during the course of the day, through which they heard, at various intervals, a randomly generated series of bleeps. They were instructed that when they heard a bleep they were immediately to 'freeze' what was passing through their consciousness at that exact moment and then make a note of it, before elaborating on it later in a follow-up interview. Although frequency varied widely, all normal (as opposed to schizophrenic) subjects reported 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. (The majority of subjects also reported the occurrence of visual images and emotional feelings—on between 0% and 50% of occasions sampled in each case). Think about it: more than half of the total set of moments that go to make up someone's conscious waking life is occupied with inner speech—that is well nigh continuous!

from it a mental model that might be imagistically expressed—and makes that available to all the various central / conceptual modules, including emotional and desire-generating systems. These then process that proposition as input, drawing inferences and generating emotional reactions, as appropriate. The result is a new cognitive context for the generation of yet another mentally rehearsed utterance, and so on.

What we have, then, is an explanation of the stimulus-independence of so much of human thought and behavior. For the initial utterance-rehearsals needn't be caused in any very direct way by stimuli impinging from outside. And once the cycle of inner speech has started, it can continue under its own momentum, with rehearsed utterances causing cognitive activity, which either causes further utterances directly, or changes the cognitive / emotional landscape against which another utterance can be generated and rehearsed.

2.3 A Problem: The Unlimited Character of Language

So far so good. But isn't there the following important difference between mental rehearsal of actions generally and mental rehearsal of utterances? In the case of physical actions, one can imagine that there might be a finite database of action schemata. And then the process of rehearsal can begin with the activation of one of these existing schemata, perhaps primed by features of the perceptual context, or activated by a heuristic like 'Take the Last'. But in the case of language there can't be a finite database of *utterance* schemata, since there is no end to the number of utterances that any competent speaker is capable of. So in this case the schemata will have to consist of utterance *components*—mostly words and phrases, but perhaps also some frequently used sentences, such as, 'What should I do next?' or, 'What am I doing wrong?' So the process of utterance rehearsal can't always begin with the activation of an existing utterance-schema. Rather, it looks as if a thought-to-be-uttered will often *first* have to be formulated, after which the utterance can be built by the language-production system combining and activating the appropriate sequence of action schemata.

This is, indeed, a significant difference between speech-actions and (some) others. (I shall return to consider some exceptions in a moment.) But there are a number of things that can be said in reply. One is that even if we confine ourselves to the standard speech-production model (Levelt, 1989)—in which utterance-generation always begins with a thought-to-be-uttered—we can still explain the stimulus-independence of much of human cognition. Granted, the initial utterance in a cycle of inner speech might be caused in a feed-forward manner by conceptual modules operating on perceptual input and competing to make their outputs available to the language production system. But as we

shall see in Section 5, the language system may have the power to combine some of these into a single utterance, whose content will therefore be different from the content of any single thought currently being entertained. And one might expect that when an utterance is rehearsed (even one that is a direct encoding of an existing thought) a whole new set of inferences and emotional reactions might result once that thought has been globally broadcast and received as input by the full range of conceptual modules. In which case, even if the first utterance-rehearsal in a cycle of inner speech is stimulus-dependent, the cycle will rapidly take on a life of its own, including utterance-rehearsals that are quite remote in content from the original stimulus.[13]

FN:13

It is also important to realize that there are domains besides that of language where we have an unlimited (potentially infinite) behavioral repertoire. Think of music and dance, for example. In each of these cases performers will have a repertoire of basic actions that they can perform. (A chord or sequence of chords played on the piano, for instance.) But there is no end to the ways in which these can be combined and recombined to make further actions of the same general type. Quite how such creative abilities are to be explained is the subject of the next chapter. But one might think that new combinations might sometimes be tried out randomly, constrained by the current context and previously performed actions, or perhaps guided by abstract resemblances to previous successful combinations. Someone improvising on the piano, for example, might at a given point select the next chord to be played at random, constrained only by whatever musical conventions are being held in place. Or they might select a chord similar to one that served well in a similar musical context recently, only in a different key.

2.4 Imagination and Episodic Memory

Thus far in this section I have focused on the way in which activations of action schemata can generate sequences of imagery, most commonly visual or auditory. But imagery surely isn't caused *only* by feedback from activated action schemata. Think how the scent of a particular flower might evoke a vivid visual image of my lover's face, or of how mention of Paris in the springtime might call up an image of us walking hand in hand through Saint Germain in the sunshine. It doesn't seem at all plausible that either of these images should be caused by an activated action schema (and in the first case, it isn't

[13] Another thing that can be said is that utterance generation and rehearsal may result from thoughts that are only weakly related to the initial stimuli, or that utilize heuristics for generating sentences that don't encode *any* prior thought. These ideas will be discussed extensively in Chapter 5, in the context of my treatment of creativity.

even clear what the relevant schema would be: surely there is no such thing as the looking-at-my-lover's-face action schema). And yet each is only loosely connected with the initial stimulus, and might give rise to a train of emotional reactions, images, and further episodic memories that could eventually take me a *long* way away from the initial stimulus.

There is reason to think that there are at least two distinct routes to the causation of a visual image, in fact (see Figure 4.4). One is the action-schema rehearsal route, discussed earlier in this section. This is argued by Kosslyn (1994) to be mechanism through which we rotate and transform images—we do so by first imagining the movements that might cause such a transformation. But another is the concept-activation route, deploying back-projecting neural pathways from temporal cortex to visual area V1, for example. These are used in normal perception to query ambiguous or degraded input, helping in the process of object-recognition (Kosslyn, 1994). In normal vision multiple concepts might be partially activated by a given visual input, and these would then be used to generate images, in an attempt to determine a 'best match'. But this same system can also be deployed 'off-line', creating images that are unrelated to current visual stimuli.

What happens when an episodic memory is formed is that a number of different things get bound together (Baddeley et al., 2002). Aspects of current experience, together with emotional reactions and beliefs about current circumstances or likely consequences—realized in different brain systems and produced by a variety of modules—get linked together and stored. Thereafter, activation of any part of this complex can serve also to activate the remainder. Perhaps perception of something similar to the imagistic aspect of the memory serves to evoke the surrounding beliefs and emotions; or perhaps a remark containing some of the crucial concepts (PARIS, SPRINGTIME) serves to evoke both the relevant beliefs and to reactivate the relevant image or sequence of images.

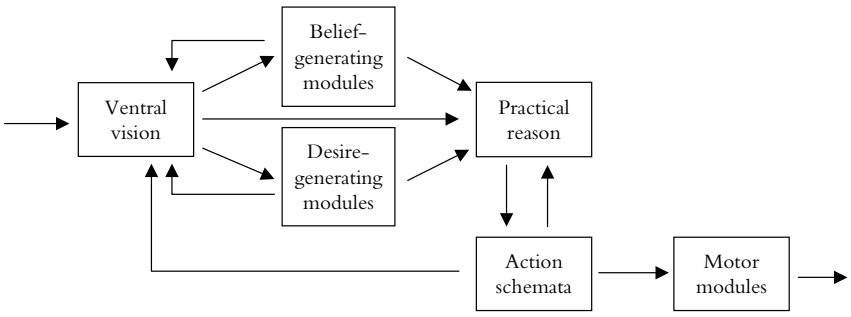

Figure 4.4. Two types of route to the generation of visual imagery

When an episode is recalled, imagery (normally visual) of the original scene will be evoked. These images are then processed by the visual system in the usual way—perhaps being elaborated and 'filled out' in the light of the subject's background beliefs and expectations—and made available to the conceptual modules once again. These in turn can generate further beliefs and further elaborations of the image, which may be stored as part of the episodic memory thereafter. (This is how memories can be elaborated and 'constructed' over time with each revisiting.)

Humans have, of course, come to enjoy evoking episodic memories for their own sake, and for the sake of the emotional reactions and rewards that come with them. Hence we have developed a variety of methods for calling up memories in ourselves and others, from the social sharing of memories in speech, through diaries, snapshots, videos, and other sorts of memorabilia. And to some extent then, episodic remembering is 'stimulus bound'. But oftentimes an aspect of the episodic memory—perhaps an element in the image that gets generated, or perhaps one of the reactivated beliefs about the circumstances—can activate yet another episodic memory, and so on. The result can be a chain of remembering that continues for some time, most of the contents of which can be unrelated to current stimuli.

Although humans now evoke memories for their own sake, that was presumably not their original function. Much more likely is that by recalling details of a previous event, one can learn something of relevance to current goals. By recalling what happened on a previous occasion of the present sort, I may be better able to figure out what I should do in the circumstances. And by recalling previous occasions in which I have interacted with someone, I may be helped in deciding whether or not to trust him now. And so on. The important point for our purposes, however, is that there is no reason to think that episodic memories are responsible for the distinctive creativity of the human mind, which will form the topic of the next chapter. For that role, the generation of action schemata seems a much more likely possibility.

2.5 Rounding up

In summary of this section, then, the stimulus-independence of human cognitive processes can be explained in two rather different ways. One is in terms of the operations of a basic capacity that we share with (some) animals, namely the capacity for mental rehearsal of action. This capacity is greatly extended by our increased, distinctively human, conceptual and knowledge-generating capacities; and especially by the development of human language, which makes possible cycles of linguistic activity in inner speech. And the other part of the explanation is in terms of episodic memory (perhaps also shared with animals;

Morris, 2002; Clayton et al., 2002), where the retrieval process gives rise to images of various sorts, which can in turn spark further episodic memories, and so on. Since all of the systems involved in both of the above sorts of process can be modules (in our weak sense of that term), there is no threat here to massively modular conceptions of the human mind.

3 Language as Content-Integrator

The hypothesis that I now want to explore and defend through the next three sections of this chapter is that natural language may be what enables us to solve the problem of content-flexibility. Versions of this hypothesis have been previously proposed by Carruthers (1996, 1998a, 2002a), by Mithen (1996), and by Spelke and colleagues (Hermer and Spelke, 1996; Hermer-Vazquez et al., 1999). I shall now sketch the thesis itself, showing how it addresses the problem of content-flexibility, before discussing (in Section 4) the experimental evidence that is alleged to support a limited version of it. In Section 5 I shall consider some alternatives and challenges. And then in Section 6 I shall turn to the problem of flexibility of reasoning-process, showing how natural language might play an important part in addressing this, as well. Section 7 will then be devoted to some clarifications and comparisons with other related proposals.

Recall from Chapter 3 that the likely shape of a language faculty, and its position within the architecture of the mind, is as depicted in Figure 3.4. It consists of distinct comprehension and production sub-systems, each of which can draw on a common database of phonological and syntactic rules, a common lexicon, and so forth. The function of the comprehension sub-system is to receive and analyze representations of natural language sentences (whether spoken, written, or signed) and—working in conjunction with other systems, both attentional and inferential—to build from that input a representation of the intended message. The latter is made available to the various belief-generating and desire-generating modules, which then get to work on that propositional content (presumably expressed in some sort of compositionally structured Mentalese or 'language of thought'),[14] evaluating or drawing further inferences, as appropriate.

FN:14

[14] As we saw in Chapter 3.4, one plausible suggestion is that the output of the comprehension sub-system is in the form of so-called 'mental models'—that is, non-sentential, quasi-imagistic, representations of the salient features of a situation being thought about (Johnson-Laird, 1983). But because mental models are compositionally structured, they nevertheless count as tokens of Mentalese. See the discussion in Chapter 1.6.

The function of the production sub-system, in contrast, is to receive input from the various central / conceptual modules (belief-generating modules, desire-generating modules, and the practical-reasoning modules), encoding the propositional content received from those systems into a representation of a natural language sentence, which can then be passed to the motor systems for expression in speech, writing, or Sign. The inputs to the language production module will be in the form of some sort of Mentalese (not necessarily the same format or code for each of the various modules in question), and the outputs will be Mentalese representations of the phonology of natural language sentences. (Representations of natural language syntax will be utilized in the interim, helping with the mapping from input to output.)

The production sub-system is ideally *positioned*, then, to conjoin contents from all of the various central / conceptual modules, since it receives input from each of them, and since it has the capacity to convert that input into a representation of a natural language sentence. But two questions immediately arise. The first is *why* the language faculty should have acquired the capacity to *conjoin* and *integrate* contents, as opposed to expressing each sequentially. The second is *how* the conjoining process is supposed to be effected.

The first of these questions is relatively easy to answer. For *speed* of utterance is one of the important design-considerations constraining the evolution of the language faculty. Whatever the precise adaptive forces that shaped the evolution of language—whether it was the mutually adaptive exchange of information (Pinker, 1994), gossip and social manipulation (Dunbar, 1996), or perhaps even sexual display (Miller, 2000)—it will be true that the faster you can frame a thought or sequence of thoughts into language and express it, the better. And certainly the speed of language production is really quite remarkably fast. (This was one of the considerations that led Fodor, 1983, to argue for the modular nature of the language faculty.) In which case, whenever the production system receives two propositional representations from different conceptual modules concerning the same event, object, or circumstance, there would have been considerable utility in being able to express those propositions in a single sentence, rather than in two separate ones.

For example, suppose that the subject is charged with describing the location of an object in a rectangular room with one colored wall. And suppose that there are distinct geometric and object-property modules that respectively deliver the Mentalese representations, THE TOY IS IN A CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE RIGHT, and, THE TOY IS NEAR A RED WALL. Then instead of having to utter two sentences separately, 'The toy is in a corner with a long wall on the left and a short wall on the right', and, 'The toy is near a red wall', the subject can just say, much more succinctly

(and hence much more swiftly), 'The toy is in a corner with a long wall on the left and a short *red* wall on the right.'

As for how this conjoining is supposed to take place, a reasonable hypothesis is that the abstract and recursive nature of natural language syntax is one crucial determinant. Two points are suggestive of how sentences deriving from two or more distinct modules might be combined into a single module-integrative one. The first is that natural language syntax allows for multiple embedding of adjectives and phrases. Thus one can have, 'The toy is in a corner with the *long* wall on the left', 'The toy is in a corner with the *long straight* wall on the left', 'The toy is in a corner with the *long straight white* wall on the left', and so on. So there are already 'slots' into which additional adjectives—such as 'red'—can be inserted.

The second point is that the reference of terms like 'the wall', 'the toy', and so on will need to be secured by some sort of indexing to the contents of current perception or recent memory. (This will be necessary when fixing the interpretation of a pronoun in an exchange of sentences with someone, for example.)[15] In which case it looks as though it wouldn't be too complex a matter for the language production system to take two sentences sharing a number of references like this, and to combine them into one sentence by inserting adjectives from one into open adjective-slots in the other. The language faculty just has to take the two sentences, 'The toy_1 is in a corner with a long $wall_1$ on the left and a short $wall_2$ on the right', and, 'The toy_1 is near a red $wall_2$' and use them to generate the sentence, 'The toy is in a corner with a long wall on the left and a short red wall on the right.'

FN:15

Notice that the integrative role of the language module, on this account, depends upon it having the capacity for certain kinds of inference. Specifically, it must be capable of taking two sentences in so-called 'Logical Form' (LF), constructed from the outputs of two distinct conceptual modules, and of combining them appropriately into a single LF representation. But it might be felt that such a claim is highly implausible, and that it is in conflict with the views of most contemporary linguists. I don't believe that either of these claims is correct, however.

It is important to see that what is in question is *not* the existence of a general-purpose inference engine located within the language faculty. For

[15] Consider an exchange in which someone says to me, 'The toy belongs to Mary.' My language comprehension system has to figure out (in cooperation with other systems, no doubt) the intended referents of both 'The toy' and 'Mary' in the context. If the thought that I then formulate as the basis for my reply has the content, *Mary plays with the toy often*, then my language production system will have to index the phrases 'Mary' and 'The toy' in such a way as to display their co-reference with the equivalent phrases in the original utterance, in order that it can be determined that it is appropriate for me to say in response, '*She* plays with *it* often', with pronouns substituted in place of the original noun phrases.

indeed, such a claim is not only intrinsically unbelievable (as well as being inconsistent with the 'one function / one module' generalization of Chapter 1), but certainly wouldn't be believed by any working linguist. However, just about every linguist *does* think (with the possible exception of Fodor and those closely influenced by Fodor) that *some* inferences are valid, and are known to be valid, purely in virtue of our linguistic competence. Most linguists believe, for example, that the inference from, 'John ran quickly', to, 'John ran' is endorsed in virtue of semantic competence (Parsons, 1990); and many would claim that such competence is embodied in the language faculty. In contemporary parlance, this amounts to saying that the inference is made by the language module transforming sentences of LF.

Admittedly, on some approaches to natural language semantics the sorts of powers and transformations envisaged at the semantic level aren't plausibly attributed to the language faculty, but would rather belong to some centralized (and pretty powerful) thought capability. This is especially true of semantics done in the tradition of Montague (1974), which presupposes a capacity to abstract arbitrarily complex concepts by deletion of components from complete thoughts, replacing those components with variables. But there are also forms of semantic theory that are much more closely integrated with Chomskian approaches to syntax, as articulated by Higginbotham (1985). Just such an account is worked out by Pietroski (2005), according to which the basic format of semantic inference is that of covert quantification over events together with conjunction introduction and reduction. Thus on this sort of approach, 'John ran quickly', really has the form, $\exists e$ *(e is a running & e is by John & e is quick)*. And then the inference to, 'John ran' is just a simple instance of conjunction elimination. On such an account, then, it is far from implausible that certain limited forms of inference (notably conjunction introduction and conjunction elimination, among others) should be handled internally within the language faculty, in transformations of LF sentences.

4 The Reorientation Data

I have suggested that if the mind were massively modular, then we should expect there to be some limits on the ways in which concepts can be combined and integrated with one another. (This prediction will be strengthened still further by some of the considerations to be adduced in Section 5.) Yet we have reason to think that there are no such limits on the flexibility of the human mind. So we have a problem: how can such flexibility of content arise in a massively modular mind? I have been suggesting that there are reasons

of a general sort for thinking that it is the language production module that performs such a role, initially in the service of speech efficiency. But Spelke and colleagues have claimed to find direct evidence in support of just this conclusion, at least in one limited domain (Hermer and Spelke, 1996; Hermer-Vazquez et al., 1999; Shusterman and Spelke, 2005). The present section will examine and discuss their argument.

4.1 The Data

The story begins with an earlier discovery of a geometrical module in rats (Cheng, 1986), already briefly discussed in Chapter 2.3. A rat disoriented in a rectangular space will rely exclusively upon geometric information when attempting to reorient itself. When its task is to search for food that it had previously seen hidden in one of the corners, it will search equally often in the geometrically equivalent corners, ignoring all other cues. One of the walls can be distinctively patterned, or distinctively scented; but the rat ignores these cues (which it is nevertheless perfectly capable of using when searching in other circumstances), and relies only upon the geometry of the space.[16] Rats therefore fail in the task roughly fifty percent of the time.

FN:16

Now admittedly, it doesn't follow from these data that rats *cannot* integrate geometric with object-property information; nor does it follow that they *don't* sometimes do so for other purposes. For all that the data show, it may be that there are links between the geometric module and the object-property module, which can lead to thoughts in other circumstances that combine concepts from both domains. But the data do at least show that in conditions of disorientation, it is only geometric information that is relied upon by the practical-reasoning system when the latter seeks to know the location of a desired target.

It should also be noted, similarly, that the finding that some other species (notably chickens and monkeys) *can* solve these sorts of reorientation tasks (Vallortigara et al., 1990; Gouteux et al., 2001) doesn't demonstrate that the members of these species are *integrating* geometric and landmark information. For the tasks can be solved by accessing the two types of information *sequentially*, first using object-property information (e.g. the location of the one red wall) to reorient, before using geometric information to guide the final stages of search. The difference between monkeys and rats may lie, not in their

[16] This makes perfectly good ecological / evolutionary sense. For in the rat's natural environment, overall geometrical symmetries in the landscape are extremely rare, and geometrical properties generally change only slowly with time; whereas object-properties of color, scent-markings, and so on will change with the weather and seasons. So a strong preference to orient by geometrical properties rather than by object-properties is just what one might predict.

powers of conjoining module-specific information, but rather in the heuristic rules that their practical-reasoning systems deploy when searching for a target object in conditions of disorientation. The monkeys may deploy the rule, 'When disoriented (and there is no directional beacon available, like a distant line of hills or the position of the sun), seek for a familiar landmark object first, and then use the local geometry', whereas the rats may use the converse ordering, 'When disoriented (and there is no directional beacon available), use the local geometry first, and then seek for familiar objects.' Indeed, this is the most parsimonious explanation of the data.[17]

FN:17

Hermer and Spelke (1996) made the startling discovery that young human children are like rats, rather than like monkeys, in this respect. The child is led into a rectangular room consisting of three white walls and one red wall, and is shown a toy being hidden in one of the corners. The child is then blindfolded and turned around until disoriented. Then the blindfold is removed, and the child is instructed to find the toy. Provided that the room is small enough that the child isn't cued to treat the red wall as a directional beacon (Leamonth et al., 2001; Shusterman and Spelke, 2005), then the child behaves just as a rat would: searching equally often in either of the two geometrically equivalent corners, and ignoring the information provided by the one red wall.

Human adults can solve these tasks, as can children older than about six or seven years. Hermer and Spelke (1996) examined the factors that predict success. It turns out that a capacity to succeed in these tasks isn't directly correlated with age, non-verbal IQ, verbal working-memory capacity, vocabulary size, or comprehension of spatial vocabulary. The only significant predictor of success in these tasks that could be discovered was the spontaneous use of spatial vocabulary conjoined with object-properties (e.g. 'It is left of the red one'). And in a follow-up study Shusterman and Spelke (2005) demonstrated that the connection is a causal one. Children who are given training in the use of 'left' and 'right', and who succeed in mastering the meanings of those terms, are much more likely to succeed when tested in a version of the reorientation task a week later.

Hermer-Vazquez et al. (1999) showed, further, that the performance of adults in the reorientation tasks is severely disrupted by occupying the resources

[17] It is tempting to seek an adaptationist explanation of these species differences. Open-country dwellers such as rats and pre-linguistic humans may have an innate predisposition to rely only on geometric information when disoriented because such information alone will almost always provide a unique solution (given that rectangular rooms don't normally occur in nature!). Forest dwellers such as chickens and monkeys, in contrast, will have an innate predisposition to seek for landmark information first, only using geometric information to navigate in relation to a known landmark, because geometric information is of limited usefulness in a forest—the geometry is just too complex to be useful in individuating a place in the absence of a landmark such as a well-known fruit tree.

of the language production module. If subjects are required to 'shadow' speech while undertaking the tasks (repeating out loud what they hear someone saying through a pair of headphones), then their performance collapses to that of younger children and rats—they, too, rely exclusively on geometric information, ignoring the information provided by the red wall. If subjects are required to 'shadow' a complex rhythm, in contrast (tapping out with their hand the rhythm played to them through their headphones), their performance isn't disrupted. So the conclusion from this, together with the childhood studies, is that it is natural language (specifically spatial language) that enables older children and adults to succeed in orientation tasks requiring them to utilize both geometric and object-property information.

4.2 Explaining the Data

So far so good. The data are quite convincing in demonstrating that it is rehearsals of natural language sentences that somehow enable older humans to solve these reorientation tasks. But do they show that the role of language is to enable the conjoining of geometric with object-property information, thus integrating the outputs of two distinct conceptual modules? Unfortunately, they do not. One salient fact is this. If the attention of younger children is drawn explicitly to the significance of the red wall (e.g. by the experimenter saying, 'Look, I'm hiding the toy by the *red* wall'), then they will succeed, despite lacking productive use of the language of 'left' and 'right' (Shusterman and Spelke, 2005). By pragmatically informing young children of the importance of the red wall, they can be cued to reorient to the red wall first, thereafter utilizing geometric information to complete their search, and following the same successful ordering of the task as do monkeys and chickens. And this (rather than a module-integrating function) may be the role that language plays in enabling older children and adults to solve the reorientation tasks, too.

Another salient fact is that adults report on the basis of introspection that the kind of sentence they rehearse when solving these tasks is, 'It is left of the red wall', rather than the more unwieldy, 'It is in the corner with a long wall on the left and a short red wall on the right.' But the former (as opposed to the latter) don't encode geometric information. The description, 'left of the red wall', combines *spatial* information with object-property information (as does, '*near* the red wall', of course). But it doesn't combine *geometric* information with object-property information; a sentence of the more unwieldy sort would be needed for that. So the role of language in enabling adult success can't be that it enables people to combine the outputs of a geometric module and an object-property module (either for the first time, or in circumstances in which such contents wouldn't otherwise be combined).

I suggest that the existing data are best explained as follows. Young children entering a reorientation task might well try out for themselves, as an aid to solving it, the natural language description, 'It is near the red wall.' But they would easily see that such a description doesn't carry adequate information, since it doesn't tell them *where* in relation to the red wall the object is. (Of course *they* don't know that they have a geometrical module, and that if they could only get themselves oriented towards red, that module would deliver the solution for them.) And they therefore discard it (i.e. make no attempt to rehearse it). But if the experimenter tells them, in effect, to look for the object near the red wall, then they rehearse the relevant sentence, and this enables them to succeed, overriding their natural disposition to rely upon geometric information first. Adults and older children, in contrast, who possess the language of 'left' and 'right', can try out the description, 'It is left of / right of the red wall', and will see that this encodes all of the information that they need to enable a solution.[18] They can then rehearse it (provided that their language production module isn't preoccupied with concurrent speech-shadowing), and can then use that sentence to guide their search-behavior once the blindfold is removed.

FN:18

But how, exactly, does mental rehearsal of a sentence guide search behavior? How does it lead to the determination of a novel action (orienting towards the red wall)? The best explanation parallels the one that we sketched in Chapter 2.8 for the role of mental rehearsal in practical reasoning generally. By mentally rehearsing, 'It is left of the red wall', the content of that sentence is extracted by the language comprehension system and globally broadcast to belief-generating and desire-generating modules, and to practical reason. The latter, on accepting a representation of the target object as being to the left of the red wall, can easily put together an action schema for the retrieval of the object. When this in turn is mentally rehearsed, an image of the subject successfully retrieving the target object is generated and broadcast, which when received by the motivational systems causes the subject to feel some satisfaction. And this in turn ratchets up the desirability of executing the action schema in question.

Although the existing data don't support the view that it is language that enables the outputs of a geometric module and an object-property module to be combined, it is important to see that they *do* nevertheless demonstrate a significant cognitive role for natural language in these tasks. For it is the rehearsal

[18] How does language enable people to succeed when the target object is placed in one of the other two corners? For (since there are *three* white walls), the description, 'Left of the white wall' doesn't uniquely identify a place. But in fact the *short* white wall is the only pragmatically salient white wall. For if the target object had been placed in either of the corners where a long white wall adjoins the red one, then the position could have been identified by means of the description, 'Left of red' or, 'Right of red'; and subjects know this.

of natural language sentences that enables human subjects to overcome their natural disposition to reorient on the basis of geometric information first. In effect, the role of natural language, here, is best assimilated to the way in which language enables flexibility of reasoning *process* (rather than content-flexibility), to be discussed in Section 6. And this is still an important result.

Moreover, I hypothesize that it ought to be fairly easy to devise a version of the reorientation tasks in which language *would* enable people to succeed by combining geometric information with object-property information. For notice that, when the tasks are conducted in a rectangular room, there is no single lexical item that means, 'corner with a long wall on the left and a short wall on the right'. The descriptor, 'left of the red wall' is therefore a great deal more convenient to use. But suppose that the shape of the room were, not rectangular, but rather rhomboid, in the form of a squashed parallelogram. Then two of the geometrically equivalent corners would be acute-angled, and two would be obtuse. And then the description, 'It is in the acute corner near red' would contain not only all the information necessary for success, but would do so in an acceptably compact form. And such a description *does* combine geometric with object-property information. We might predict that at least some adults would deploy such a sentence to enable them to succeed in the tasks. And we might predict that young children who don't yet have the vocabulary of 'left' and 'right', but who are given training with the terms 'acute' and 'obtuse' (or more accessibly, perhaps, with the phrases 'pointy corner' and 'wide corner'), might thereby be enabled to succeed.

5 Alternative Theories of Content Flexibility

Although the existing experimental data don't directly support a content-integration account of the cognitive role of language, it seems highly likely that such data could be obtained with the right experimental manipulation. And I have argued that there are in any case general theoretical reasons for taking seriously the idea that it is language that enables us to combine the outputs of some different belief-modules. (Recall that a great deal of content-integration will already be taking place by virtue of the wiring connections that exist between some belief modules and some other belief modules.) Now in the present section I shall consider two alternatives to the account of content-integration sketched in Section 3. One is that the module doing the work of content-integration isn't the language faculty, but rather the mind-reading system. The other is that there might be a special-purpose content-integrating mechanism downstream of the central / conceptual modules, positioned between them and

the practical-reasoning system. In the course of this discussion yet further reasons for taking seriously the content-integration role of language will emerge.

5.1 Meta-Module or Language Module?

Atran (2002b) presents the following challenge to the above account: Why should we think that it is language that does inter-modular integration, rather than some sort of meta-representational faculty—a theory of mind mechanism (ToMM) or a mind-reading mechanism? Atran agrees with the modularist framework adopted here; and he agrees that domain-general flexible cognition somehow has to be built out of modular components. But he sees nothing yet to discriminate between my proposal and the idea previously advanced by Sperber (1996) that it is the mind-reading module that has the power to combine the outputs of all the others.

As we saw in Chapter 1, it is perfectly plausible that some modular systems might routinely exploit the resources of other systems independently of language, querying those systems for needed information. And this will surely be the case for a mind-reading mechanism. Indeed, just such an account of the operations of the mind-reading faculty has recently been offered, and ably elaborated and defended, by Nichols and Stich (2003), as we saw in Chapter 3.3. So I fully accept that the system in question can access some of the contents generated by these other systems, no matter whether they be concerned with mates, financial institutions, falling stones, or whatever. The point is just that the mind-reading system itself should be incapable of drawing any further inferences from these contents, except those mandated by its own inferential principles.

So the mind-reading module will be able to go from, 'John has *seen* Mary with a basket of red tomatoes', to, 'John probably *knows* that Mary has a basket of red tomatoes' (in virtue of the mind-reading principle, 'seeing leads to knowing'). But the mind-reading system itself won't be able to infer, 'John knows that Mary has some *ripe* tomatoes.' To get that inference, it will have to send out a query elsewhere, and get the response, 'Red tomatoes are ripe', or the response, 'Mary has a basket of red tomatoes $\supset$ Mary has a basket of ripe tomatoes', and then rely on the mind-reading principle, 'People know the obvious consequences of other things that they know.'

The main point here is one of task-analysis, combined with the form of 'one function / one module' generalization defended in Chapter 1. Attributing mental states to other people on the basis of behavioral cues, and/or predicting people's behavior from mental states previously attributed to them, is what the mind-reading system is primarily about. Combining concepts and propositions, and drawing arbitrary inferences from them, would seem to be a distinct set

of functions entirely. In which case we should expect that the latter functions will be carried out in one or more distinct modules. For there is no reason to think that they will come 'for free' with mind-reading functions. In contrast, there *is* good reason to think that the content-combining functions *will* come for free with the language module, given the constraint of speed of sentence-production.

Another (related) difficulty for the Sperber / Atran proposal is to explain why their proposed meta-representational inter-modular architecture should have evolved. For the main business of the mind-reading faculty is presumably to predict and explain the behavior of conspecifics. This capacity would only have required the construction of inter-modular thoughts if others were *already* entertaining such thoughts. Otherwise attributions of module-specific thoughts alone would have done perfectly well. (The same point is valid, of course, even if the primary purpose of the mind-reading faculty were the introspective ascription of thoughts to oneself. For there would only be a point in self-ascribing an inter-modular thought to oneself if one were already capable of entertaining such a thought, prior to the operations of the mind-reading faculty.)

In the case of language, in contrast, the demands of swift and efficient communication would have created a significant selection pressure for inter-modular integration, allowing the outputs of distinct central modules concerning the same object or event to be combined into a single spoken sentence. So instead of saying separately, 'The object is near a short wall', and, 'The object is near a red wall', one can say much more succinctly, 'The object is near a short red wall' (given that the short wall in question *is* the red wall, of course).

It might be replied that the pressure for the mind-reading system to integrate contents across modules could have come, not from the demands of predicting and explaining the behavior of oneself and others, but rather from the benefits that such integration can bring for other areas of activity (such as solving the reorientation problem). This is possible. After all, it is common enough in biology that a system initially selected for one purpose will be co-opted and used for another. And it might be claimed that the mind-reading system would be ideally placed to play the integrative role, receiving information from all the other central modules, and providing outputs to practical reasoning.

But actually, it is hard to see how one could get from here to anything resembling the full flexibility of human cognition. For there is no reason to suppose that the mind-reading system would have been set up in such a way as to provide its output *as input to* the other central modules (as opposed to *querying* those modules for needed information). In which case there would be no scope for *cycles* of reasoning activity, with the mind-reading system combining the

outputs from central modules and then feeding the conjoined content back to them, harnessing their resources for purposes of further reasoning. In contrast, since language is both an output *and* an input module, it is well positioned for just this role, as I argued in Section 3.

In conclusion of this sub-section, then, I claim (on both general theoretical and evolutionary grounds) that language is a much more plausible candidate for integrating the contents of other central modules than is the mind-reading system.[19]

FN:19

5.2 A Special-Purpose Content-Integrator?

Each of the proposals considered so far maintains that content-integration is carried out by systems that initially evolved for other purposes (communication, in the case of language; explanation and prediction of other people's actions, in the case of mind-reading). The obvious competitor hypothesis is that content-integration is undertaken by a special-purpose module of the mind that was designed to do just that: integrate contents. This system would have to be located downstream of the various central / conceptual modules, from which it would receive its input. And it would need to make its output available to the language production system, and perhaps also to practical reason (see Figure 4.5).

It might be said that this model has an advantage over my language-based one. This is that language production can here operate entirely along classical lines. The language production system will receive a complete integrated thought of some sort—THE TOY IS IN AN ACUTE-ANGLED CORNER NEAR A RED WALL, as it might be—and it will encode that thought into speech. No inferences will need to be drawn, and no content-conjoining needs to take place within the language faculty. It is doubtful, however, whether this is a

[19] In addition, suppose we were satisfied that the dual-task data collected by Spelke and colleagues (Hermer-Vazquez et al., 1999) are about content *integration* rather than content *sequencing*; or suppose that we could successfully obtain such data. That would then support my own proposal quite strongly, when matched against the Sperber / Atran one. For if subjects fail at the task in the speech-shadowing condition (but not in the rhythm-shadowing condition) because they cannot then integrate geometrical and landmark information, it is very hard to see how it can be a disruption to mind-reading that is responsible for this failure. For there is no reason to think that shadowing of speech should involve the resources of the mind-reading module. Admittedly, a good case can be made for saying that normal speech comprehension and production will implicate a sub-system of the mind-reading module, at least—where that sub-system is charged with figuring out the conditions for *relevance* in communication (Sperber and Wilson, 2002). So normal instances of speaking and comprehending would surely disrupt any other concomitant task that involves mind-reading (and so any task requiring inter-modular integration, on the hypothesis that it is the mind-reading module that performs this role). But there is no reason to think that this should be true of speech *shadowing*. For in this case there is no requirement of comprehension, and no attempt at communication.

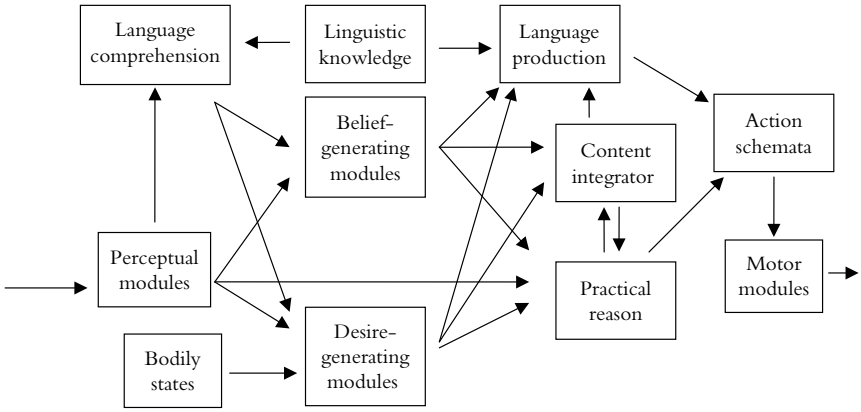

Figure 4.5. The existence of a distinct content integrator

very powerful consideration in support of the model. For if many linguists are already inclined to believe, on other grounds, that the language faculty has limited inferential abilities, and is capable of combining words and phrases about the same subject matter, then we shall already have all the materials necessary for the language-based account of content-integration to operate.

If we are to believe in a special-purpose content-integrator, moreover, then we had better be provided with some account of the evolutionary pressures that might have led to its existence. What could these have been? Not the demands of swift and efficient communication, presumably, since these would have been much more likely to give rise to the minor alterations in the language-production system required for it to take on the content-integration role, as we saw in Section 3. Nor, I think, can the existence of a content-integrator be explained by the benefits that might accrue to intra-modular processing, deriving from integrating the outputs of two or more modules that both feed their output into a given (third) module. For this would have created pressure for an adaptation in that latter module itself, rather than for the creation of a whole new system to perform the task.

The explanation of content integration via language that we provided in Section 3 had to do with speed of communication of thoughts deriving from distinct modular systems. And then it might be said that pressures of speed, not on communication, but rather on practical reasoning, might have given rise to a special-purpose content-integrator. For in that case, instead of having to receive two or more distinct thoughts in sequence, the practical-reason system could receive just a single integrated thought. But this would have required a corresponding increase in the complexity of the algorithms being operated by

the practical-reason system. In fact, it would have had to develop some way of 'looking into' an atomic proposition to discern and utilize its distinctive integrated content. And besides the obvious costs in terms of computational complexity, these operations would also take up additional time, of course; so there is unlikely to have been any overall gain in speed.

It seems to me, then, that the costs of feeding integrated contents to the practical-reason system would have more than outweighed any benefits. In the case of language, in contrast, the initial pressure for changes in the systems that consume the newly integrated linguistic contents would have lain outside the head of the speaker, in the language comprehension systems of his or her interlocutors. One can easily imagine a sort of evolutionary 'arms race' taking place, here, in which the benefits deriving from increases in speed of communication drive small changes in the language production system, enabling distinct sentences to be integrated with one another before utterance, with corresponding pressure on the language comprehension system to upgrade itself as a result.

The only other way that I can imagine the idea of a distinct content-integration module being fleshed out would be to postulate that this is actually the work of a logic module, of the sort that Sperber (1996) has suggested might exist. This would be a module that responds just to the *forms* of the propositions that it receives as input, looking for opportunities to derive some logical consequences from them. So if the module receives two representations, P, and, IF P THEN Q, as input then it generates the consequence, Q, as output; and so forth.

There may well be such a logic module. Certainly the capacity to derive some of the simpler logical consequences of what you believe would be a useful capacity to have. But it can't do the work required of it in the present context, as we noted in Section 1.6. This is obviously the case if the system in question deals only in propositional logic, since such logics treat propositions as 'atoms', whose contents can get *combined* with one another (e.g. by conjunction introduction) but never *integrated*. And even if the system implements some sort of predicate logic with identity, this still doesn't have the resources to effect the sort of inference required. In the simplest case, we want to be able to integrate the contents of two propositions like, THAT_1 IS A SHORT WALL, and, THAT_1 IS A RED WALL, to form the representation, THAT_1 IS A SHORT RED WALL. But predicate logic doesn't have the resources to construct complex predicates like, 'short red wall'. This seems much more likely to be the work of a language faculty, whose business is precisely to build up complex sentences and phrases from their parts.

There is one final consideration that counts against the existence of any sort of special-purpose content-integrator, and in favor of our language-based

account. This arises out of the following fact. There is nothing to guarantee that the outputs of all belief-generating modules possess the same representational format; and it seems quite likely, indeed, that those outputs *aren't* all in the same format. For example, the output of the geometrical module might be in some analog quasi-spatial, quasi-perceptual, format; whereas the output of the object-property system might have a digitalized sentence-like format. So the evolution of a content-conjoiner module would at the same time have had to be the evolution of a system that can interface with systems that employ a variety of representational formats. This means that there would have to have been some especially strong pressure for the existence of such a module.

Something similar is true of the language system, of course. It too, on the production side, would need to be capable of receiving input from all the different belief-generating and desire-generating modules, whose outputs might differ from one another in their representational format. But this doesn't set any special puzzle for a content-integrating account of the role of language, since interfacing with conceptual modules is something that a language faculty would need to do anyway. And it would be natural language itself that would provide the 'common format' in which the outputs of the different conceptual modules could be integrated. We don't need to postulate any extra selection pressure on the language faculty to enable it to integrate the contents that it encodes; the abstract and recursive nature of natural-language syntax will see to that. Whereas, in contrast, we *do* need to postulate an extra-strong selection pressure for the existence of a content-conjoining system, to enable it to interface with central modules.

In the case of language, there exist plausible accounts of the adaptive forces that might have led to its evolution, as we have seen. And some of these accounts postulate that the evolutionary process might have been gradual and incremental, precisely in respect of the number of other systems with which the language faculty can interface. (Dunbar, 1996, for example, supposes that language was initially for communication of *social* information. So presumably, on this account, the interfaces with the various social modules—mind-reading, cheater-detection, and so forth—would have been built first; with interfaces to other systems only being added later.) But what would have been the selection pressure for a special-purpose content-integrator? As we saw earlier, it is very hard to discern one.

5.3 Conclusion of Sections 3, 4, and 5

I have sketched an account of the way in which a language module might be responsible for the seemingly unlimited content-flexibility of the human mind, and I have contrasted this account with alternatives. While the experimental

evidence that has been alleged to support such an account currently falls short of doing so, it seems quite likely that this lacuna can be filled. And there are a number of general considerations that support the proposal over the available competitors.

6 Inner Speech and the Flexibility of Reasoning

Thus far we have sketched modularist, language-based, explanations of both the stimulus-independence of so much of human cognition and the content-flexibility of human thought. But what of the flexibility of human reasoning *processes*—the acquisition and maintenance of new patterns of thinking, and new rules of reasoning? Can this be explained within the framework outlined above? I believe that it can. But first I need to explain and elaborate the so-called 'two systems' account of human reasoning processes.

6.1 Two Reasoning Systems and Mental Rehearsal

Virtually all of the scientists who study human reasoning and the pervasive fallacies that so often occur in human reasoning have converged on some or other version of a two-systems theory (Evans and Over, 1996; Stanovich, 1999; Kahneman, 2002). System 1 is really a *collection* of systems, arranged in parallel. These systems are supposed, for the most part, to be universal (common to all members of the species), to be evolutionarily ancient, and to operate swiftly and unconsciously. Moreover, their processing algorithms are either immutable, or subject to their own idiosyncratic trajectory of learning and change—at any rate, explicit instruction has little impact on their operations. In the context of the present discussion, then, they can be identified with the set of central / conceptual modules.

System 2, in contrast, is supposed to operate linearly (rather than in parallel), and to be slower and characteristically conscious in its operations. But it can override or pre-empt the results of System 1. And its algorithms are much more mutable, and are more easily influenced by explicit teaching of various sorts. System 2 is also much more subject to individual variation. Thus Stanovich (1999) shows that variability in success in the various standard reasoning tasks (which are thought to require the operation of System 2) correlates highly with IQ, and hence also with *g*. And even when IQ is factored out, it correlates highly with certain measures of variable cognitive *style* (such as a disposition to be reflective, and a capacity to distance oneself from one's own intuitive views).

As I have already remarked, within the context of massively modular models of the human mind we can identify the parallel System 1 processes with

the operations of a set of central / conceptual modules. And we can then understand System 2 as supervening on the activity of those systems, realized in cycles of mentally rehearsed action in general, and inner speech in particular. We will return to consider in some detail how such cycles can give rise to novel beliefs and novel actions in Chapters 6.7 and 7.4 respectively. But for the moment it should be noted that speech is an *activity*, of course. And like any other activity, sequences of action can be learned and practiced, and can improve as a result of explicit instruction. And one can also reflect on those sequences, and form beliefs about their appropriateness. Let me elaborate.

Consider a simple action like lifting an object from the floor and placing it on a table-high surface. There is a more or less intuitive and natural sequence of actions that most people will perform in a case of this sort. One will approach the object, bending from the waist (perhaps with some bending of the knees, depending on how high the graspable surface of the object rides above the floor), grab it with arms fairly straight, and then lift by straightening one's back while at the same time bending one's arms at the elbow. This works very effectively in a wide range of situations. But as many people now know, it can be a recipe for back-injury when the object to be lifted is heavy, or when many lighter objects have to be lifted in a repetitive sequence. In such cases, the proper sequence of action is to approach the object, go almost into a squatting position in front of it, bending one's knees fully while keeping one's back straight, grasping the object with arms bent at the elbow, and then lifting by straightening one's knees while keeping one's arms fairly immobile.

This novel sequence of action can of course be learned by imitation, or as a result of explicit verbal instruction; and it can be practiced until it becomes smooth and natural—indeed, *habitual*. And it is possible to reflect on it or alternatives to it, in the sort of way that physical trainers and athletes often do—debating, either externally with others or inwardly with oneself, which sequence of act-components would best realize one's overall goals. And having debated, and reached a conclusion, it is possible to train oneself to act accordingly.

Likewise then, suppose that I am faced with a version of the Wason card-selection task, and am asked which of the four cards I should turn over to tell whether or not the target statement, 'If P then Q', is true.[20] Like so many other

FN:20

[20] In these now-famous experiments, subjects are confronted with four cards on which propositions are inscribed, of the form: P, $\sim P$, Q, $\sim Q$. They are told that on the back of each of the $P/\sim P$ cards will be inscribed either Q or $\sim Q$; and that on the back of each of the $Q/\sim Q$ cards will be inscribed either P or $\sim P$; and that these four cards correspond to the truth-value combinations of P and Q in the envisaged situation. And their task is to decide which of the four cards they need to turn over to decide whether or not the statement, 'If P then Q' is true.

people, I turn over just the *P* card, relying on my swift-and-intuitive System 1 cognitive processes. But then I take a course in propositional logic, in which the instructor informs me that a conditional is only false in the case where the antecedent is true and the consequent false. 'Remember', the instructor might say, 'in order to evaluate the truth of a conditional you need to look for a case where the antecedent is true and the consequent false.' Then, when faced with a Wason selection task on some later occasion, I take a moment to reflect, ask myself (in inner speech), 'What do I have to do in order to evaluate a conditional?' The global broadcast of the content of this question enables me to recall what I was told, and hence I correctly turn over both the *P* and *not-Q* cards. And of course it is possible to reflect on *patterns* of sentences used in reasoning, too, trying to develop rules by means of which to reason better, as logicians and scientific methodologists have traditionally done.

Consider, in addition, how we might explain the main elements of cognitive style that correlate so well with success in difficult (System-2-involving) reasoning tasks (Stanovich, 1999). The disposition to be reflective before giving an answer or reaching a decision maps nicely into our model. This will be the disposition to engage in explicit conscious thinking in inner speech or other forms of imagery before responding. And the capacity to distance oneself from one's own intuitive views, too, finds a ready explanation on our model. For this can be identified with a readiness to entertain alternative suppositions, formulated and mentally rehearsed in inner speech or other forms of imagery. And that each of these elements of cognitive style should correlate with success in the sort of reasoning that requires access to learned rules is exactly what would be predicted, on our account.[21]

FN:21

The proposal being made here, then, is that in addition to a set of swift and unconscious modular thinking and reasoning processes (System 1), humans also engage in a kind of reasoning that is slower (realized in *cycles* of System 1 activity), that is conscious (by virtue of the global broadcasting of sensory representations of utterances in inner speech and other action rehearsals), and that is to a significant degree language-based. The latter (System 2) is realized in sequences of action-schema activations (often rehearsals of natural language

[21] An interesting further prediction is that people who are introverted should be disproportionately successful at difficult System-2-involving tasks, and should thus be over-represented in populations of gifted individuals. (I owe this observation to Chris Pyne.) For introverts have a well known tendency to be both self-stimulating and introspective (presumably occupying their time in cycles of visual imagery and inner speech). And indeed, a recent survey of research on this topic finds that introverts are to be found significantly more often amongst gifted students in comparison to control populations (Sak, 2004).

utterances), with the sequences taking place (sometimes) in accordance with learned rules and inferential procedures.

6.2 System 2 as Mental Rehearsal: For and Against

We will explore this proposal in greater detail in future chapters. But some additional support for it can be found in the extended defense of the 'think aloud' protocol for research into human cognitive processing mounted by Ericsson and Simon (1993). They argue that in connection with a wide range of cognitive tasks (most of which we would now categorize as 'System 2'), reliable evidence of how people actually solve these problems can be gathered by requiring subjects to 'think out loud', in overt speech, while they tackle them. They emphasize that it is crucial not to ask people to comment on, or describe, how they are solving the problem. For provision of such a meta-commentary will demonstrably interfere with the normal performance of the task in question, and will thus provide unreliable evidence of how such tasks are normally undertaken. But if subjects are encouraged to 'speak their thoughts', then the thoughts that they articulate will actually map nicely onto one or another of the various *possible* ways of solving the problem, as gleaned from task-analysis. And the resulting account will also be supported by other indirect measures, such as the length of time that the task takes.

These facts are easily and neatly explained if what people are doing when they attempt to solve problems in a System 2 manner (or at least when the tackle the sorts of System 2 problems tested by Ericsson and Simon) is that they engage in cycles of verbal rehearsal in inner speech. For then all that the think-aloud protocol does is remove the inhibitory process that would normally prevent the activated motor schemas, which are used to generate the sentences of inner speech, from emerging in overt action.

There is also a natural *objection* to the proposal being made here, however. For I am suggesting both that System 2 processes play an important role in explaining individual variations in *g*, and that such processes are realized in mental rehearsals of action, including mental rehearsals of speech action, in 'inner speech'. These claims would seem to imply that people whose language system is damaged or destroyed should always perform poorly on standard psychometric tests. But some aphasics can show spared performance in such tests (Kertesz, 1988; Varley, 1998). And likewise, there are sub-populations of children with Specific Language Impairment (SLI) whose general intelligence is within the normal range (van der Lely, 2005). Let me reply to these points separately.

It is very doubtful whether my proposal concerning the role of language in System 2 processes entails that children with SLI should have lower IQs than normal. This is because the deficits involved (within the relevant sub-population, at least) are relatively subtle ones. Children in this group make frequent errors in their use of actives and passives, in agreement, and in tense. They might say, for example, 'He fall to the ground', rather than, 'He fell to the ground.' But it is quite unclear why this should lead to any failures in the child's own reasoning. (It certainly shouldn't interfere with the content-conjoining role of language outlined in Section 3.) For when the child mentally rehearses such a sentence when problem solving, there will generally be ample cues to enable the child's language comprehension system to extract a past-tensed thought from the present-tensed rehearsed sentence. If there are any effects here, they will be subtle time-to-completion ones, which wouldn't show up in an IQ score.

As for IQ in aphasia, there are two points to stress. One is that my view is *not* that System 2 reasoning per se is language-dependent. Rather, it is that System 2 is dependent on mental rehearsals generally, with language-rehearsal being just one (very important) form amongst others. And the second point is that *verbal* psychometric tests, of course, cannot even be administered to aphasics. So it might be the case that the reasoning underlying successful performance in *those* tests is verbally mediated. And the fact that an aphasic subject can have preserved abilities to reason with Raven's Matrices, for example, or in tasks that require pictures to be ordered to make a coherent narrative, should not be surprising, on my view. For the relevant rehearsals can be conducted in visual imagery rather than speech.

6.3 Thinking as Acting

Notice that the account sketched here has the resources to explain our common-sense belief that thinking is often an *activity*, and is something that we *do* (being at least partly under our intentional control), rather than something that merely *happens* within us (Frankish, 2004).[22] This belief might initially seem to be problematic in the context of a massively modular conception of the human mind. For if the mind consists entirely of modular special-purpose

FN:22

[22] I should stress that I don't mean to be claiming that *all* conscious thinking results from the activation and rehearsal of action schemata. On the contrary, some visual images (e.g. of a window shattering from an impending impact) are produced from the output of our belief modules, utilizing the back-projecting neural pathways in the visual system that are also used in object-recognition (Kosslyn, 1994). And visual and other images can also be passively evoked when an episodic memory gets activated by some cue. What I shall claim in Chapter 5, however, is that all *creative* thinking results from the activation of action schemata. So all conscious thinking that is *creative* is a form of acting.

systems (each of which just goes ahead and does its own processing-job automatically), and if all thinking has to be realized in the operations of such systems, then how *could* thinking be active rather than passive? But now the solution to this puzzle is easy to see, if System 2 thinking consists largely in sequences of action-rehearsal (including utterance-rehearsals). For these are a species of *action*—not to be assimilated to consciously planned activities undertaken to achieve certain goals in the light of one's beliefs, perhaps; but at least similar to routine bodily movements and sequences of movement like stretching, driving a car over a familiar route, or stepping around an obstacle in one's path.

Since System 2 thinking is a species of acting, thinking *skills* can be acquired via any of the variety of mechanisms through which behavioral skills are normally learned. One such mechanism is explicit instruction. People can *tell* me what actions and sequences of action to perform in the service of a given goal. As noted above, a logic teacher can tell me what steps to take in order to evaluate the truth of a conditional, just as an experienced kayak maker might tell a novice how to prepare and shape the materials for use in the frame. And in both cases the beliefs thus acquired can be recalled at the point of need, and used to guide novel instances of the activity in question (evaluating a conditional, building a kayak).

Another way of acquiring skills, of course, is imitation. Many complex skills are learned through forms of apprenticeship in which both explicit instruction and trial-and-error learning may play some role, but in which novices also spend extended periods of time working alongside, and observing and imitating, experienced practitioners. Where the skills in question are thinking skills, this can only happen when the teacher chooses to 'think aloud' for the benefit of the novice. Much of what happens in scientific lab meetings should perhaps be seen in this light, where problems are reasoned through publicly in the presence of undergraduate and graduate research assistants. And likewise, many college lecturers see themselves not only as imparting information to their students, but also as *exemplifying* in their lectures the patterns of thinking and reasoning that the students are to acquire. Certainly I have always looked at my own lecturing in this way, as providing my students with an example of how one might think through a problem. I intend them to imitate the forms, if not the contents, of my thinking.

Whatever the mode of their initial acquisition, thinking skills, like all other skills, can become smoother, swifter, and less error-prone with practice. Much of what takes place in mathematics classrooms can be understood in this light. Through rote learning of their times-tables, children acquire and gradually render habitual a series of action-sequences. And when learning how to do

addition sums and division sums, they acquire behavioral procedures that are initially slow and halting, and are conducted overtly (often on paper), but which with time and practice can become both habitual and internalized, in inner speech or sequences of visual imagery.[23]

FN:23

The present account of many forms of conscious thinking—in terms of the activation, rehearsal, and global broadcast of action schemata—can also mesh nicely with our best accounts of *disorders* of thinking, such as frequently occur in schizophrenia (Frith, 1992; Campbell, 1999). Patients with schizophrenia often complain that their thoughts are not their own. On the contrary, they claim, their thoughts are being *inserted* into their minds by members of an alien race, or by government agents, say. And likewise they often claim that their overt actions are not their own, either. A patient might complain, for example, that when he picks up a comb and runs it through his hair, it is not he who controls his hand but some outside force. This commonality between delusions of thought-insertion and delusions of behavior-control is easily explained given an account of (System 2) thinking as a species of acting.

According to Frith et al. (2000), part of what has gone wrong in patients with schizophrenia involves damage to the system that utilizes efferent copies of motor commands to construct a sensory representation of the intended outcome, normally used for purposes of comparison, self-monitoring, and self-correction. Because of this, schizophrenics don't feel that their own actions (including the thought-actions that issue in inner speech and sequences of visual and other imagery) are their own, even though they are intentional, and even though the outcomes are the ones that were intended. And just as this explanation predicts, it turns out that schizophrenic patients are incapable of using *mental practice* to improve performance through internally generated feedback, and they can't make swift corrections to their behavior in the light of sensory feedback, either (Frith et al., 2000). And also as predicted, they have

[23] Evidence consistent with these suggestions is presented by Spelke and Tsivkin (2001), who conducted three bilingual arithmetic training experiments. In one, bilingual Russian–English college students were taught new numerical operations; in another, they were taught new arithmetic equations; and in the third, they were taught new geographical and historical facts involving both numerical and non-numerical information. After learning a set of items in each of their two languages, subjects were tested for knowledge of those items in both languages, as well as tested on new problems. In all three studies subjects retrieved information about exact numbers more effectively in the language in which they were trained on that information, and they solved trained problems more effectively than new ones. In contrast, subjects retrieved information about approximate numbers and about non-numerical (geographical or historical) facts with equal ease in their two languages, and their training on approximate number facts generalized to new facts of the same type. These results suggest that one or another natural language is implicated in thought about exact numbers, but not when representing approximate numerosity (a capacity shared with other animals).

difficulty in distinguishing between experiences (such as a tickle on their hand) that are self-produced and those that are other-produced.

6.4 Norms for Thinking

Recall from Chapter 3.7 that one of the likely modules that is specific to the human mind is a system for learning, storing, and reasoning with *norms*, as well as for generating intrinsic motivations towards compliance with them. Most norms are rules governing behavior. They tell us what we must, must not, or are permitted to *do*. Hence if System 2 thinking is a species of behaving, as I have been arguing that it is, then it is easy to understand how the very same normative system can come to govern much of human thinking and reasoning as well. As a result, there will be at least three different *kinds* of ways in which System 2 thinking / behaving can be motivated. Two have already been mentioned above. One is that the activation of a particular goal habitually calls up a certain action-sequence in its own service. Another is that the subject believes that a certain action or sequence of actions is a reliable way of achieving a given goal. But now the third is that the subject believes that a certain type of action is mandated / required (or forbidden) in the context, and thus feels intrinsically motivated to think / behave in the appropriate way.

Consistent with this prediction, people do seem to be intrinsically motivated to entertain or to avoid certain types of thought or sequence of thought. Thus if someone finds himself thinking that P and thinking that $P \supset Q$, then he will feel *compelled*, in consequence, to think that Q. And if someone finds herself thinking that P and also thinking that $\sim P$, then she will feel herself *obligated* to eliminate one or other of those two thoughts. (And in circumstances where a contradiction isn't easily eradicated—such as arguably occurs in classical electrodynamics (see Frisch, 2005)—people will take steps to ensure that the contradiction doesn't very often surface in consciousness, but rather remains, for the most part, covert. This may be because contradictory thoughts, like normative breaches generally, make us feel uncomfortable.) The explanation is that the norms module has acquired a rule requiring sequences of thought / action that take the form of *modus ponens*, as well as a rule requiring the avoidance of contradiction, and is generating intrinsic motivations accordingly.

The action-theory of System 2 thinking, combined with a norms module for constraining and guiding action, can thus explain how so much of our conscious thinking should take place in a sea of normative beliefs, or should seem to occur within what some have called 'the space of reasons' (McDowell, 1994). But it is important to distinguish the account being advanced here

from the often-defended—but in my view radically mistaken—claims that norms of reasoning are *constitutive* of thinking as an activity, and that correct ascriptions of thoughts to people are constrained and partly constituted by norms of rationality (Davidson, 1973; Dennett, 1987; McDowell, 1994; and many other philosophical writers). It is worth spending a few paragraphs to explain the difference.

On my account it is *beliefs about* norms, rather than norms themselves, that do the explanatory work. Representations involving modal concepts like MUST or MUSTN'T are stored in a special-purpose belief-box, attached to a reasoning system that is continually on the lookout for circumstances that might lead to a match with the non-normative content of those representations. When one is found, the system generates an intrinsic motivation towards performing the action described by that content (in the case of 'must'), or against acting (in the case of 'must not'). Whether the represented norms are *true* or *correct* is quite another matter. And the resulting account is fully consistent with a naturalistic, purely causal, account of thinking and believing.

According to the account endorsed by some philosophers, in contrast, the very notions of *thought* and *belief* are themselves intrinsically normative. Thinking itself is said to be an inherently *rational* process, such that any deviations from ideals of rationality make it more difficult for us to conceive of the agent as a thinker at all. And attributions of thought to an agent have to be constrained by the rational norms (such as 'avoid inconsistency') that govern all thinking.[24] In consequence, such philosophers can't offer a fully naturalistic account of what thought itself *is*. Davidson (1970) thus maintains that while each *token* thought is a physical state of the brain which has causes and effects, thoughts themselves (as types) play no causal role in the world. (This is his so-called, 'anomalous monism'.) And likewise Dennett (1987) maintains that beliefs and thoughts have no reality independent of our practices of thought-ascription, in which we adopt what Dennett calls 'the intentional stance' for purposes of predictive convenience.

I am not wanting to claim that beliefs about norms of reasoning play no role whatever in our attributions of thoughts to others, of course. Quite the contrary. For the way in which we generally go about attributing to someone the thoughts that they will arrive at by reasoning from a given starting point is to utilize *our own* reasoning processes in a partial 'simulation' of the reasoning of the other person, as we suggested in Chapter 3.3. And then if our own System 2 reasoning involves beliefs about norms of reasoning, the outcome

[24] For a particularly elegant and powerful critique of this sort of view, see Cherniak (1986). See also Nichols and Stich (2003).

will be a function of what we take those norms to be. But to reiterate, it is our beliefs about norms of reason, not norms themselves, that are to some degree constitutive of our practices of thought-ascription. And thoughts themselves can be fully naturalistic entities that are *not* normatively constituted.

7 What the Thesis is and Isn't

The picture that I have been developing in this chapter is that a number of the kinds of flexibility that are distinctive of human thought processes can be partially explained by supposing that those processes are conducted in inner speech, recruiting both the resources of the language module and the resources of a wide range of central / conceptual modules (as well as the motor-control modules used to build and streamline increasingly sophisticated action schemata). In the present section I shall attempt to clarify these ideas further, by contrasting them with others with which they might naturally be confused, and by comparing them with some other familiar models.

7.1 Thinking in Language?

According to the views that I have been developing, natural language has important cognitive functions (in addition to its obvious communicative ones). And one might gloss this by saying that some of our thinking is conducted in rehearsals of natural language sentences. But does this really mean that some System 2 processes are conducted *in* language, and/or that natural language sentences serve as the *vehicles* for our System 2 thoughts? And if so, can these ideas actually be made sense of? For there are familiar and devastating objections to the claim that we (English people) think in English (Jackendoff, 1997; Pinker, 1997; Pylyshyn, 2003).[25]

FN:25

For example, one objection is that English sentences are almost always radically incomplete encodings of the intended thought. If someone says, 'The fridge is empty', they will have some particular standard of emptiness, and some particular fridge, in mind. (Do they mean that there isn't enough in the fridge to make a meal, so that the statement might count as true even though the fridge contains some lettuce leaves and a bottle of milk? Do they mean that the fridge is now ready for cleaning, in which case what they say will be false if it contains bottles of milk, but true if it contains only crumbs of bread and

[25] We might also enquire how, in more detail, we are to explain how new beliefs can be arrived at in such a manner. But this question will be deferred to Chapter 6. For it is intimately connected with the question of how abductive / scientific reasoning is to be explained, as we shall see.

cheese? Or do they intend a standard consistent with the fact that it now *has* been cleaned, in which case the statement is false if the fridge contains even crumbs? Moreover, which fridge is intended as the referent of '*the* fridge'?) The intended standards and referents will have to be gathered, by inference, from the context. Since attaching a specific content to the sentence, 'The fridge is empty', requires thought, that sentence cannot by itself carry the content of the thought that it communicates.

In fact it would be highly misleading to express the view that I am developing by saying that natural language sentences are the *vehicles* of System 2 thinking. For, first of all, there aren't actually any sentences of English contained in the human mind / brain. Rather, there are (Mentalese) *representations of* English sentences, and it is these representations that do (or are involved in doing) the real work. Secondly, one should in any case say that natural language sentences (or rather representations thereof) are *implicated in* System 2 thought processes, or are essential *components of* those processes, rather than that they are the *vehicles* of those processes. Let me elaborate on each of these points.

My hypothesis is that some human thought processes involve rehearsals of natural language utterances. And for a native English speaker, of course, these will be rehearsals of sentences of English, each consisting in an activated action schema that contains a representation of the sentence in question. This action schema is used to generate a representation of what it would be like to hear (or see, in the case of Sign) the corresponding utterance. And this quasi-perceptual imagistic representation is globally broadcast and made available *inter alia* to the language comprehension system, which attaches a content to it and makes that content (as expressed in some sort of Mentalese representation, perhaps in the form of a mental model) available to the suite of central / conceptual modules. Nothing in this story requires us to talk about *sentences* of English figuring in cognition, as opposed to *representations of* English sentences. Nor does it presuppose that the computations that underlie our basic thought processes are defined over such representations.

But how can these representations of natural language utterances be constitutive of *thinking*? In answering this, it will help to introduce a simple convention. Let us use single quote marks to designate representations of natural language sentences, and line brackets to designate all other Mentalese expressions. And then imagine a case where what gets rehearsed is a representation of the natural language sentence 'P', and where the Mentalese sentence that gets constructed by the comprehension system on receipt of the representation 'P', is |Q|. (Since there is no guarantee that the comprehension process will issue in a Mentalese representation with the same content as was used to construct the natural language sentence in the first place, it is safest—here and

elsewhere—to consider examples where it doesn't.) And let us suppose for simplicity that |Q| has the content *that Q*.

Now according to a number of the hypotheses sketched above, the pairing <'P', |Q|> has further consequences in cognition—and not just *any* consequences, but those that are distinctive of thinking. One way in which this might be the case is if the representation |Q| is one that can *only* be formed (either absolutely, or in context) via the construction of an appropriate natural language sentence, as 'module-integration' accounts of the role of natural language in cognition suggest (Hermer-Vazquez et al., 1999; Carruthers, 2002a). The sentence 'P' is constructed by combining and integrating a number of other sentences—'R' and 'S', say—individually produced from the outputs of a pair of conceptual modules. When 'P' is mentally rehearsed the comprehension system extracts from it the Mentalese representation |Q|, which can then (especially if it takes the form of a mental model) be globally broadcast for all the different central modules to get to work upon, provided that any aspect of it meets their input-conditions. The result might be further thought-contents that would never have been entertained otherwise.

Another way in which pairing <'P', |Q|> might have consequences that are distinctive of thinking is if it is only by virtue of articulating the sentence 'P' in auditory imagination, and hence making its content available to the various inference-systems that exist downstream of perception and consume its products, that the subject comes to believe |Q| (i.e. comes to believe *that Q*) for the first time. (See Chapter 6.7.) The process of articulating 'P' leads to |Q| being evaluated and accepted, in a way that would not—as a matter of fact and given the circumstances—have happened otherwise. In such a case it seems perfectly sensible to say that the act of articulating 'P' is part of the process of *thinking* that leads to the generation of a new belief (the belief *that Q*).

Yet another way in which the pairing <'P', |Q|> might have consequences distinctive of thinking can be derived from the account of the dual systems of reasoning discussed in Section 6. If the agent has learned, through training, explicit instruction, or imitation, to engage in those speech / action sequences in which sentences of type Φ should always be followed by sentences of type Ψ (where 'P' belongs to type Φ and 'R' belongs to type Ψ), then the pairing <'P', |Q|> will lead to some further pairing of the form <'R', |S|> . In which case the process that leads the subject to entertain the thought *that S* will be one that constitutively involves representations of natural language sentences (in this case, representations of the sentences 'P' and 'R').[26]

FN:26

[26] The rules governing such action-sequences may sometimes be highly abstract, as are the rules of formal logic. But there is no objection to the proposal to be raised from this direction. For even bees

7.2 Weak Whorfianism?

Many philosophers and social scientists throughout the twentieth century maintained that language is the medium of all human conceptual thinking. Most often this claim has been associated with a radical empiricism about the mind, according to which virtually all human concepts and ways of thinking, and indeed many aspects of the very structure of the human mind itself, are acquired by young children from adults when they learn their native language. And it has been held that these concepts and structures will differ widely depending upon the conceptual resources and structures of the natural language in question. This mind-structuring and social-relativist view of language is still dominant in the social sciences, following the writings early in the last century of the amateur linguist Whorf (many of whose papers are collected together in his 1956). Indeed, Pinker (1994) refers to it disparagingly as 'the Standard Social Science Model' of the mind.

My views should be distinguished sharply from those of the Standard Social Science Model, of course. I maintain, on the contrary, that most of our concepts, and many of our forms of thinking and learning, are independent of, and prior to, natural language. And I maintain that much of the architecture of both human and animal minds is innately fixed, and that the mind contains many innately structured learning mechanisms. In recent decades a weaker set of Whorfian views have been explored by cognitive scientists, however. According to such views natural language doesn't *create*, but rather *sculpts* or *shapes* human cognitive process. For example, acquisition of Yucatec (as opposed to English)—in which plurals are rarely marked and many more nouns are treated grammatically as substance-terms like 'mud' and 'water'—is said to lead subjects to see similarities amongst objects on the basis of material composition rather than shape (Lucy and Gaskins, 2001). And children brought up speaking Korean (as opposed to English)—in which verbs are highly inflected and massive noun ellipsis is permissible in informal speech—are said to be much weaker at categorization tasks, but much better at means–ends tasks such as using a rake to pull a distant object towards them (Gopnik et al., 1996; Gopnik, 2001).

The basic idea behind weak Whorfianism can be expressed in terms of Slobin's (1987) idea of 'thinking for speaking'. If your language requires you to describe spatial relationships in terms of compass directions, for example, then you will continually need to pay attention to, and compute, geocentric

can learn to generate actions in accordance with the abstract rule, 'Turn right in the second chamber if it is marked in the same way as the first, turn left in the second chamber if it is marked differently from the first.' They can generalize to new forms of marking (such as new combinations of color), and also generalize across sense modalities—trained on colors, they can generalize to instances of the rule involving odors. See Giurfa et al., 2001.

spatial relations; whereas if descriptions in terms of 'left' and 'right' are the norm, then geocentric relations will barely need to be noticed. This might be expected to have an impact on the efficiency with which one set of relations is processed relative to the other, and on the ease with which they are remembered (Levinson, 1996). Likewise in respect of motion events: if you speak a language, like English, that employs an extensive and often-used vocabulary for *manner* of motion ('walk', 'stride', 'saunter', 'hop', 'skip', 'run', 'jog', 'sprint', etc.), then you will continually need to pay attention to, and encode, such properties. In languages like Spanish and Greek, in contrast, manner of motion is conveyed in an auxiliary clause ('He went into the room *at a run*'), and it often goes unexpressed altogether. One might then predict that speakers of such languages should be both slower at recognizing, and poorer at remembering, manner of motion (Slobin, 1996). This claim has been subjected to careful experimental scrutiny by Papafragou et al. (2002), however, who are unable to discover any such effects.

Levinson's claims for the effects of spatial language on spatial cognition have also been subject to a lively controversy (Levinson, 1996, 2003; Li and Gleitman, 2002; Levinson et al., 2002; Li et al., 2005). Let me pull out just one strand from this debate for brief discussion. Levinson (1996) had tested Tenejapan Mayans—who employ no terms meaning *left* and *right*—on a spatial reversal task. They were confronted with an array of four items on a desk in front of them, and told to remember the spatial ordering of three of the items. They were then rotated through 180° and walked to another table, where they were handed the three items and told to 'make them the same.' As predicted, the Mayans employed geocentric rather than egocentric coordinates when complying with the instruction, just as the hypothesis of 'thinking for speaking' would predict.

In the course of their critique, however, Li and Gleitman (2002) point out that the task is plainly ambiguous. The instruction, 'make them the same', can mean, 'lay them out similarly in respect of geocentric space', or it can mean, 'lay them out similarly in respect of egocentric space.' (And indeed, Westerners who are given these tasks will notice the ambiguity and ask for clarification.) Li et al. (2005) therefore reason that Levinson's results might reflect, not an effect of language upon thought, but rather an effect of language upon language. Since the instruction is ambiguous, subjects are presented with the problem of disambiguating it before they can respond appropriately. And since geocentric descriptions are overwhelmingly more likely in the society to which the Mayans belong, they might naturally assume that the instruction is intended geocentrically, and act accordingly. It doesn't follow that they would have had any particular difficulty in solving the task in an egocentric fashion if cued

accordingly. And for all that the experiment shows, they might routinely deploy egocentric concepts in the course of their daily lives (if not in their daily speech).

To test this, Li et al. (2005) devised a series of unambiguous spatial tasks that admit of only a single correct solution. In one of these, for example, the subjects had to match a card containing two differently sized circles to one of four cards of a similar sort, but variously oriented. Once they were familiar with the task, they were allowed to study the card at one table before being rotated 180° and walked to a second table where four cards were laid out for them to match against. But they did this under one of two conditions. In one, the card was covered and carried to the other table while they watched without its orientation relative to the Earth being changed. (This is the geocentric condition.) In the other, the card was placed in their hands and covered before they turned around through 180° to face the other table. (This is the egocentric condition.) Contrary to Levinson's predictions, the subjects did just as well or better in the egocentric condition. And when the task demands were significantly increased (as when Li et al. had subjects recall and trace out one particular path through a maze under two conditions similar to those described above), the Mayan subjects actually did significantly better in the egocentric condition (80% correct versus 35% correct).

In an earlier presentation of some of the views defended in this chapter I was concessive about the powers of different natural languages to sculpt cognition differently during development, as weak Whorfian accounts of the role of language in cognition maintain (Carruthers, 2002a). However, there is no particular *need* for me to be concessive towards the 'language sculpts cognition' approach. Each set of data will have to be examined on a case-by-case basis, of course. And at the moment the jury is still out on the question whether language sculpts cognition to any significant degree. But certainly the weak forms in which this thesis is currently being pursued are consistent with the strong modularism adopted in the present book, and also with my main theses. But no such weak Whorfian claims are supported or entailed by my views. And the empirical data are still subject to a variety of interpretations. Accordingly, this topic will now be dropped for the remainder of our discussions. It has been mentioned here only in order to contrast it with the cognitive functions of language to which I am committed (i.e. the functions of stimulus-independence, content-flexibility, and flexibility of reasoning process).

7.3 Vygotsky and Dennett

In the present section I shall contrast my views with those of Vygotsky (1961) and Dennett (1991), who each present proposals that are significantly (and in my view, unacceptably) stronger than my own.

At around the same time that Whorf was writing, Vygotsky was developing his ideas on the interrelations between language and thought, both in the course of child development and in mature human cognition. These remained largely unknown in the West until his book *Thought and Language* was first published in English (Vygotsky, 1961). This attracted significant attention, and a number of further works were translated through the 1970s and 1980s (Vygotsky, 1971, 1978; Wertsch, 1981, 1985). And some of Vygotsky's claims have obvious points of contact, as well as elements of contrast, with my own.

One of Vygotsky's ideas concerns the ways in which language deployed by adults can *scaffold* children's development, yielding what he calls a 'zone of proximal development'. He argues that what children can achieve alone and unaided isn't a true reflection of their understanding. Rather, we also need to consider what they can do when scaffolded by the instructions and suggestions of a supportive adult. And such scaffolding not only enables children to achieve with others what they would be incapable of achieving alone, but also plays a causal role in enabling children to acquire new skills and abilities. Relatedly, Vygotsky focuses on the overt speech of children, arguing that it plays an important role in problem solving, partly by serving to focus their attention and partly through repetition and rehearsal of adult guidance. And this role doesn't cease when children stop accompanying their activities with overt monologues, but just disappears inwards. He argues that in older children and adults *inner* (sub-vocal) speech serves many of the same functions.

Several of these ideas have been picked up by later investigators. For example, the self-directed verbalizations of young children during problem solving activities have been studied. One finding is that children tend to verbalize more when tasks are more difficult, and that children who verbalize more often are more successful in their problem solving (Diaz and Berk, 1992). The thesis that language plays such scaffolding roles in human cognition isn't (or shouldn't be) controversial. But in Vygotsky's own work it goes along with a conception of the mind as socially constructed, developing in plastic ways in interactions with elements of the surrounding culture, guided and supported by adult members of that culture.

These stronger views—like the similar constructionist views of Whorf (1956)—are inconsistent with the thesis of massive mental modularity being defended in this book, at least if interpreted in any robust form. But a restricted version of them can survive as an account of the development of System 2 thinking and reasoning. For as we have seen, such thinking is to a significant extent dependent upon inner speech, which both supervenes on and recruits the activity of underlying modules. And as we have also seen, many of the patterns of activity that take place within System 2 are learned from others (via

both instruction and imitation), as well as being guided by socially acquired norms of reasoning.

Let us turn now to Dennett (1991). He famously argues that human cognitive powers were utterly transformed following the appearance of natural language, as the mind became colonized by *memes* (ideas, or concepts, which are transmitted, retained and selected in a manner supposedly analogous to genes; see Dawkins, 1976). Prior to the evolution of language, on this picture, the mind was a bundle of distributed connectionist processors. These conferred on early hominids some degree of flexibility and intelligence, but were quite limited in their computational powers. The arrival of language then meant that a whole new—serial and compositionally structured—cognitive architecture could be programmed into the system.

This is what Dennett calls the *Joycean machine* (named after James Joyce's 'stream of consciousness' writing). The idea is that there is a highest-level processor that runs on a stream of natural-language representations, utilizing learned connections between ideas, and patterns of reasoning acquired in and through the acquisition of linguistic memes. On this account, then, the concept-wielding mind is a kind of social construction, brought into existence through the absorption of memes from the surrounding culture. And on this view, the conceptual mind is both dependent upon, and constitutively involves, natural language.[27]

FN:27

Here, too, there is much that I can agree with, as well as disagree with. Of course I can agree that cycles of inner speech both sustain and partially constitute our System 2 thought processes. And I can also agree that much of the activity of System 2 depends upon beliefs and norms of rationality that have been acquired from the surrounding culture, and that System 2 operations exemplify patterns of activity that have been learned from other people, either by instruction or imitation. But I shall also claim that most concepts and structured thought processes are independent of language, involved in the operations of numerous System 1 modules. And even the role of socially acquired memes, within System 2, should look quite different when seen through the lens of massive modularity. Or so I shall now briefly argue.

Like most others who use the notion of a 'meme' as an explanatory construct, Dennett (1991) thinks of memes as passively acquired items of cultural

[27] Admittedly, what Dennett will actually *say* is that animals and pre-linguistic hominids are capable of conceptual thought, and engage in much intelligent thinking. But this is because he is not (in my sense) a realist about thoughts. (See Chapter 2.1.) On the contrary, he thinks that there is nothing more to thinking than engaging in behavior that is *interpretable as* thinking. Yet he commits himself to saying that it is only with the advent of natural language that you get a kind of thinking that involves discrete, structured, semantically evaluable, causally effective states—that is, thoughts *realistically construed*.

information. But as Sperber (2000) points out, very little cultural learning is a mere matter of absorbing and retaining information. On the contrary, most socially communicated information needs to be *reconstructed* through processes of (module-based) inference of various sorts. And this means that the transmission process will be heavily biased by the underlying modular architecture. These ideas are especially nicely illustrated in the work of Boyer (2001), who shows that the seeming cacophony of religious ideas in the myriad religions of the world are actually organized around the central modular domains of psychology, living beings, non-living physical things (e.g. mountains), and artifacts. Almost all religious beliefs concern things that are drawn from one or other of these domains, but with properties that violate some of the central assumptions of the module in question (for example, a stone statue that can listen to prayers). Such beliefs thereby combine maximum memorability (from the violated expectation) with maximum inferential potential (from the normal operations of the module in question).

7.4 Clark and Jackendoff

By way of yet further clarification, in the present section I shall consider some views of the role of language in cognition that are significantly weaker than those that I have been defending.

Clark (1998) draws attention to the many ways in which natural language is used to scaffold human cognition, defending a conception of language as a cognitive *tool*. (Chomsky, too, has argued for an account of this sort. See his 1975, ch. 2.) Such instrumental uses of language range from the writing of shopping lists and post-it notes, to the mental rehearsal of remembered instructions and mnemonics, to the performance of complex arithmetic calculations on pieces of paper. According to this view—which Clark labels 'the supra-communicative conception of language'—certain *extended* processes of thinking and reasoning constitutively involve natural language. The idea is that language gets used, not just for communication, but also to augment human cognitive powers (especially by enhancing memory).

Thus by writing an idea down, for example, I can off-load the demands on memory, presenting myself with an object of further leisured reflection. And by performing arithmetic calculations on a piece of paper, I may be able to handle computational tasks that would otherwise be too much for me (and my working memory). In similar fashion, the suggestion is that *inner* speech serves to enhance memory, since it is now well-established that the powers of human memory systems can be greatly extended by association (Baddeley, 1990). Inner speech may thus facilitate complex trains of reasoning, by enabling

us to hold their component parts in mind in a way that would otherwise be impossible (Varley, 1998).

Notice that on this supra-communicative account, the involvement of language in thought only arises when we focus on a process of thinking or reasoning *extended over time*. So far as any given individual (token) thought goes, the account can (and does) buy into the standard conception of language as a mere input–output device. It maintains that there is a neural episode that carries the content of the thought in question, where an episode of that type can exist in the absence of any natural language sentence and can have a causal role distinctive of the thought, but which in the case in question causes the production of a natural language representation. This representation can then have further benefits for the system of the sort that Clark explores (off-loading or enhancing memory).

According to the account of the cognitive role of language presented in this chapter and the one following, in contrast, a particular tokening of an inner sentence is (sometimes) an inseparable part of the mental episode that carries the content of the thought-token in question. So there is often no neural or mental event at the time that can exist distinct from that sentence, which can occupy a causal role distinctive of that sort of thought, and which carries the content in question; and so language is actually involved in (certain types of) cognition, even when our focus is on individual (token) acts of thinking.

Jackendoff (1996, 1997), likewise, puts forward an account of the role of language in cognition that is weaker than that being defended in this book. (See also Carruthers, 1996, for presentation and defense of some similar ideas.) He suggests, in particular, that inner speech serves to focus our conscious *attention* on our thoughts and thought processes. As a result, two sorts of further cognitive effect tend to ensue. One is that the thought or thought process in question is subjected to more detailed processing, by virtue of being 'anchored' in working memory through expression in inner speech. This might lead, for example, to the development of a more comprehensive plan of action, or to the generation of further thoughts that are consequences of the one under consideration. The other effect is that the thought or thought process becomes available to meta-cognitive awareness, enabling it to be questioned, criticized, and improved.

These proposed cognitive roles for language are quite plausible, and are fully consistent with the ideas presented in the present chapter. I have suggested that the mechanism by means of which inner speech achieves a focusing of attention upon thought is via the global broadcast of sensory representations of natural language utterances. And many of the further effects of those broadcasts result from the contents of the utterances in question being received as input

by the myriad conceptual modules, creating cycles of modular processing. (Jackendoff, too, presents his ideas within a modularist framework, albeit one that isn't quite so massively so.) And likewise I have suggested that, because speech is a form of action, inner speech can enable our thought processes to become subject to norms of rationality and truth.

The ideas that I am defending go further than this, however. Both here and in Chapter 5 I suggest that language can enable thought-contents to be formulated for the very first time, and/or be entertained in circumstances where they would not otherwise occur. In the present chapter I have argued that language plays a role in conjoining otherwise disjoint module-produced concepts. And in the chapter following I shall argue that it plays a role in the creative generation of wholly novel thoughts, too. In addition, I have argued that sequences of sentences in inner speech occur as they do not *just* because of the way in which modular processes get to work on the contents of each (as Jackendoff, 1997, suggests). Rather, because speech is an activity, speech-sequences of a given type can be learned as an acquired skill, through instruction or imitation.

7.5 Baddeley's Working Memory

In this section I shall compare and contrast the model of human cognitive architecture sketched in the present chapter with the account of the working-memory system developed over a number of years by Baddeley and colleagues (Baddeley and Hitch, 1974; Baddeley, 1986, 1990, 1993; Gathercole and Baddeley, 1993). Both theories postulate short-term working-memory systems intimately linked to such cognitive functions as planning, reasoning, and conscious awareness; and both assign a role to imagistic representations of language within the systems described.

Baddeley has proposed that the working-memory system consists of a central executive and two specialized slave-systems, the *visuo-spatial sketchpad* and the *phonological loop*. The relationships between them are represented in Figure 4.6 (adapted from Gathercole and Baddeley, 1993). The central executive controls the flow of information within the system as a whole, and is charged with such functions as action-planning, retrieval of information from long-term memory, and conscious control of action. The executive also allocates inputs to the visuo-spatial sketchpad and phonological loop, which are employed for spatial reasoning tasks and language-related tasks respectively. Since the central executive must presumably have access to linguistic knowledge, if it is to be able to generate linguistic inputs to the phonological loop, this model could easily be presented in such a way as to resemble fairly closely the model represented earlier in Figures 4.3 and 4.4.

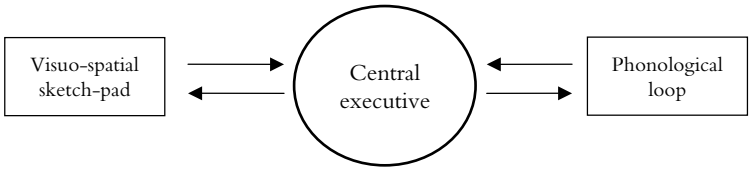

Figure 4.6. Working memory

One difference from my model concerns the special-purpose nature of the phonological loop. In particular, Baddeley seems to think of it as *essentially* a phonological system. In contrast, my model proposes that we can, in principle, entertain linguistically formulated thoughts through the imaginative use of any language-related sense-modality. In normal individuals, no doubt, such thinking involves auditory, or perhaps articulatory (kinesthetic) imagination (or both). But in the case of those whose only native language is some form of Sign, my account predicts that their linguistic thinking will involve the manipulation of *visual* (or kinesthetic) images. And perhaps some ordinary thinkers, too, sometimes employ visual images (in this case of written language) in their language-based thinking.

One empirical prediction of my model, then, is that exactly the sorts of interference-effects that have been used to explore the properties of the phonological loop in normal subjects would be found in the visual (or perhaps the kinesthetic, gestural) modality for deaf subjects whose native language is a form of Sign. Another prediction is that aphasics or other brain-damaged patients who have lost the phonological component of working memory should be able to recover their capacity for language-based thinking by employing the resources of some other form of imagination—either kinesthetic, developing an articulatory loop, or visual, manipulating images of written sentences. For according to my model, the exact form in which linguistic information is represented in language-based thinking is plastic, and may vary from individual to individual, and within individuals over time.

Another difference between Baddeley's model of working memory, on the one hand, and my massively modular model of System 2 processes, on the other, concerns the *function* of the phonological loop—its causal role in the activity of the cognitive system as a whole. In Baddeley's account the phonological loop is employed *only* for language-based tasks—that is, only for tasks that are explicitly *about* language, or explicitly *involve* language. Thus the phonological loop is said to be involved in such tasks as memorizing sequences of letters, vocabulary acquisition, reading development, and language comprehension. But there is no suggestion that it is also involved in the planning of action,

or in other forms of reasoning about the world (rather than about language). These tasks are allotted, rather, to the central executive.[28]

FN:28

In my own model, in contrast, the phonological loop (and/or its equivalent in other sense-modalities) is involved in many forms of conscious thinking and reasoning about the world. Recall that my hypothesis is that some of our occurrent thoughts are formulated in the form of images of natural language sentences, which are then globally broadcast to a wide range of inferential systems. (See Figure 4.3.) By virtue of such broadcasting our cognitive system is able to gain access to some of its own processes of thinking, in such a way as to render them conscious.[29] And cycles of sentence production and global broadcasting make possible System 2 thinking and reasoning. The function of the phonological loop is thus much more than just to enable the mind to engage in language-involving processing tasks. It is also to enable the overall system to gain access to its own occurrent thoughts, thus facilitating cycles of such thought, as well as the sort of indefinite self-improvement that comes with self-awareness and System 2 thinking and reasoning.

FN:29

I do have to concede, of course, that there is also a need for something resembling Baddeley's central executive within my own account. For *something* must be responsible for selecting and manipulating the imaged sentences in the phonological loop, which therefore become the system's conscious occurrent thoughts (in virtue of the reflexive availability of the contents of the loop). And likewise something must be responsible for selecting and manipulating the visual images in the visuo-spatial sketchpad, which are similarly globally broadcast. But on my account, the system in question is a sort of *virtual* executive, involving the interactions of many different belief-generating systems and action-selecting systems. It isn't itself a distinct isolable system. Thus competition between modules to present their outputs as input to the language system might play a role. And as we shall see in Chapter 5, associations amongst related concepts will also be important. My hope would be that an account along the lines of the one being developed here could eventually be seen as a workable *realization* of Baddeley's model within a massively modular mental architecture.

[28] In fact Gathercole and Baddeley note in passing, following Hitch (1980), that the phonological loop may be implicated in mental arithmetic; see their 1993, p.234. But nothing further is made of this point. For evidence supporting such a view, see Spelke and Tsivkin (2001).

[29] This is true whether one endorses a *first*-order account of what makes mental events conscious (Baars, 1988, 1997; Tye, 1995, 2000) or whether one endorses a *higher*-order account (Carruthers, 2000, 2005). According to the former, conscious status is a matter of the availability of the mental events in question to systems charged with belief-formation and practical reasoning—and this is what global broadcasting achieves. But according to the latter, consciousness is a matter of availability to higher-order thought, giving us *awareness of* our thoughts. And this, too, is achieved by global broadcasting. For one of the systems to which imagistic events will be broadcast is a mind-reading faculty capable of higher-order thoughts about those events. See Chapter 3.3.

8 Conclusion

In this chapter I have articulated the main challenges that a thesis of massive mental modularity must face (albeit with 'module' taken in the weak sense defended and adopted in Chapter 1). I have distinguished several different ways in which human thought and behavior might be said to be distinctively *flexible*. And our task has been to show how these forms of flexibility can be accommodated and explained within a massively modular model of the human mind. Some of these challenges have proven relatively easy to meet—at least, on the assumption that representations of natural language sentences play a constitutive role in human cognitive processes. (These are the objections from context-flexibility, from stimulus-independence, from content-flexibility, and from the flexibility of human reasoning processes.) Other challenges are harder, and will be confronted in the chapters that follow.