PART III

# Elaborating Mechanisms

In order to really understand something such as speech, which is peculiar to man, it will be necessary to find ways of recording from single neurons from outside the skull.

—D. Hubel 1979

Since the human mind resides in the brain, we cannot be satisfied with our explanations of human thinking until we can specify the neural substrates for the elementary processes of the human symbol system. Of these connections we know next to nothing. . . . We are in a position similar to that of 19th-Century chemistry, which had developed an extensive theory of chemical combination long before that theory could be linked to the physics of atoms.

—H. Simon 1980

## INTRODUCTION

In Part 2 we explored the preliminaries to mechanistic explanation. We turn now to the process of developing and elaborating such explanations. This requires showing how an activity that is performed by a whole system is accomplished by having different components perform different functions that contribute to the task. This is where the heuristics of decomposition and localization most properly come into play. Decomposition involves establishing a division of labor according to which different tasks involved in the same overall process are identified. Localization entails a systematic and independent examination of the processes operating at the lower level and a demonstration that these processes perform the functions specified in the decomposition.

We saw in Chapter 5 that direct localization was often charged by opponents with being vacuous or spurious. The natural consequence of this charge was a rejection of mechanism and a shift to vitalistic or dualistic views. Yet those who sought at all costs to avoid vitalistic or dualistic results could not simply neglect the power and force of the objections. The direct localizations we considered in Chapter 4 *do* posit control at the lower level for the very phenomena they are meant to explain. It was for this reason, together with the fact that the functions being studied were not obviously mechanistic, that the explanations were regarded as spurious. Put another way, direct localization was viewed as *insufficiently constrained localization*. In Gall's program, for example, higher-level cognitive capacities were localized in physical organs, but the only reason he offered for attributing specific capacities to specific organs was the need to explain the capacities. This is part of what we meant to imply by calling the evidence for Gall's view *merely* correlational. Flourens saw that this was not yet an explanation, and certainly not an explanation in terms of lower-level mechanisms. Similarly, Pasteur saw that the attempt to explain fermentation in terms of catalytic processes depended on showing that fermentation could be carried out in the absence of living organisms, and his classic experiments were designed to show that it could not. If fermentation could not be performed without living organisms, then the postulated catalysts could not be simple chemical agents, and the appeal to chemical mechanisms for fermentation could offer no serious explanation. It is the lack of *independent* constraints on what functions are localizable that makes the solutions *insufficiently* constrained.

The point is a general one. We are told that the capacity for language comprehension is localized in regions posterior to the Sylvian fissure; that the cerebellum is the locus for motor control; that the left hemisphere is

specialized for "analytic" tasks (while the right is more concerned with pattern and synthesis); and that there are genes for altruism, parental care, manic depression, schizophrenia, and Alzheimer's dementia. These are all claims to localization, but they are by no means all on an equal footing with respect to either evidential or theoretical credentials. Some are well motivated. Some are not. One factor that is important for our purposes in Part III is *the character of evidence* that serves to underwrite such claims. A well-developed localizationist model depends on detailing the mechanisms, robustly characterized. It is important that we can detail the mechanisms, and it is important that the evidence renders them robust. In some cases, such mechanisms can be isolated and characterized; in others, they cannot.

Let us consider two relatively recent examples. There have been highly publicized accounts claiming to have isolated a gene for manic depression (within a small population) on a single chromosome (see Egeland et al. 1987). It has long been accepted that manic depression has a significant hereditary component. More recently, researchers have claimed to isolate genetic markers on the tip of the short arm of chromosome eleven that are common among those members of an Amish community in southeastern Pennsylvania who suffer from manic depression. As the researchers describe it, there is a linkage between these markers and "a locus conferring a strong disposition to bipolar affective disorders" (ibid., p. 784). The researchers are, of course, under no illusion that there is a single gene controlling manic depression, and they explicitly allow that other genes are involved as well as that there is heterogeneity in manic depression's genetic basis; nonetheless, they claim that there is a correlation between a specific chromosomal pattern and the presence of manic depression within that Amish community and that, within that population, this chromosomal pattern is indicative of a gene resulting in manic depression. It is important to recognize how tightly circumscribed this claim is. There is no claim to know the mechanism underlying the etiology of the disease. The data is limited to a specific population; moreover, there is obvious evidence that this correlation is limited in scope.[1] This is a clear claim to localization, though the underlying mechanisms are admittedly complex. What we wish to underscore is that the *evidence* for the claim is correlational only. Independent physical constraints are simply lacking. We do not know how the correlated chromosomes could constitute causal agents in the etiology of the disease; in fact, we do not even know *whether* the genes identified are causal agents or just simple markers.

It is interesting to compare this claim with another recent, and analogous, claim to have localized the defect responsible for Alzheimer's disease (see Barnes 1987). This latter claim suffers few limitations paralleling

those present in the research into manic depression. On the one hand there is evidence from one group of researchers (Goldgaber et al. 1987) that indicates the defect in at least one form of Alzheimer's disease—the inherited variant—is located on chromosome twenty-one. This evidence is drawn from information on four distinct family groups. Other research locates a gene coding for an amyloid protein in the proper region on chromosome twenty-one (Tanzi et al. 1987), and yet another group reports an accumulation of amyloid protein filaments in the brains of aged mammals and an unusually high deposition of this protein in Alzheimer's patients (Selkoe et al. 1987). Finally, there is clear evidence that in Alzheimer's patients there is a gene duplication not found in control subjects that would, presumably, result in an exaggerated dosage of the amyloid protein (Hyman et al. 1984). All of this is hardly conclusive and is insufficient to determine the precise etiology of Alzheimer's disease. We do, though, have solid reasons to think that amyloid proteins are involved, even if we do not yet see exactly how.

What is important for our purposes is that this case not only provides evidence of discrete localization, but also connects this localization with evidence suggestive of a mechanism—or, more precisely, with evidence suggesting that there is a determinate genetic mechanism. An impressive array of evidence converges on the same result. What is notably lacking in the study of the Amish community is any biochemical clue as to why the genes are implicated in the disease, or how they might contribute to it. The Alzheimer's case, by contrast, offers robust evidence for the localization, and some of the evidence is at least indicative of an underlying mechinism.[2] We will see that this sort of coupling, drawing on independent theories and evidence at more than one level, is important to localizationist research.

The cases to be described in Chapters 6, 7, and 8 all share one important feature not present in the cases of Part II: lower-level theories, and empirical results accessed at those lower levels, impose significant constraints on the development of the relevant theories. That is, in the terms we used in Chapter 1, the lower-level theories provide constraints that limit the search space. It turns out that these constraints are powerful. They provide information about the class of allowable mechanisms and so limit the space of mechanisms needing to be searched (much as the assumption that relations between variables were linear aided BACON in inducing Boyle's Law). The strength of these constraints—the extent to which they effectively limit the search space—varies. In Chapter 6 the constraints are relatively weak and fundamentally empirical. In Chapters 7 and 8 they are substantially stronger and have a theoretical component as well as an empirical one. We will also see that many of the constraints are modified or

abandoned altogether. As more constraints are imposed, and the task is one of simultaneously satisfying independent constraints, then it sometimes becomes necessary to compromise some in order to satisfy others.

In Chapter 6 we examine work in the neurosciences. One central case is the classic work on aphasias by Karl Wernicke toward the end of the nineteenth century, and the second is a more recent psychobiological investigation of spatial memory carried out by John O'Keefe and Lynne Nadel (1978). On the basis of systematic deficits observed in patients, Wernicke concluded that there were localized centers for a variety of associations. One important feature of Wernicke's work, in contrast to that of, say, Broca, is that it was explicitly based on neuroanatomical evidence concerning the projections of cerebral tracts. O'Keefe and Nadel construct a model more severely constrained by the empirical data. After examining these cases, we will discuss the role of localization and decomposition in the interpretation of the experimental results and show that the localizationist program in its stronger forms requires a convergence of cognitive models and neural mechanisms.

Though O'Keefe and Nadel begin with a direct localization of objective cognitive and spatial maps in the hippocampus, they take up the task of explaining how the hippocampus performs this function. This requires a shift to a lower level. Accordingly, they focus on the components of the hippocampus and not on the whole organ. This introduces lower-level constraints. They appeal to empirical data detailing different patterns of neuronal activity in different parts of the hippocampus, including both single-cell recordings and EEG patterns. They then proceed on this basis to develop a functional analysis showing how the activities occurring in different components of the hippocampus at different stages in the performance of spatial-problem-solving tasks could accomplish the overall task of spatial mapping. This is the sort of model that we refer to as a complex localization. The data concerns the behavior of individual cells and patterns in their activity. The correlation of the activity of individual cells with differences in spatial-problem solving provides a more substantial empirical basis for localizing spatial representation in the hippocampus. The decompositional analysis of the hippocampus, and demonstration of how its components could carry out the necessary processing to be the seat of the capacity to develop and use objective mental maps of spatial layouts, serves two functions. First, it provides independent, and hopefully convergent, empirical evidence vindicating the initial direct-localization claim; and second, it offers a mechanistic account of how the task is performed.

Chapters 7 and 8 examine cases imposing still stronger lower-level constraints, including both empirical and theoretical components. Chapter 7 focuses on research devoted to determining the fermentation mechanism

in living cells. As we observed in Chapter 5 fermentation was a process that seemed to elude mechanistic explanation in the nineteenth century. The major breakthrough came with Buchner's demonstration in 1897 that fermentation could be achieved in a cell-free extract. Despite his initial direct localization, researchers quickly realized that one enzyme could not accomplish the complex reaction of fermentation. Rather than developing the mechanistic explanation by shifting to a lower level and finding subcomponents that performed the function through their interaction, research remained at the same level and sought to identify a variety of components that would each play a different role in carrying out the overall task of fermentation. It was not possible to identify and distinguish enzymes physically, so researchers had to proceed functionally instead. They investigated various ways of inhibiting part of the functional pathway, and they stimulated the operation of the pathway by supplying potential intermediaries in an attempt to determine the stages in the fermentation process.

The first complex-localizationist models proposed for fermentation assumed that the process consisted principally of a linear series of reactions through which sugar was rendered into alcohol. In the language of Chapter 2, researchers assumed that the fermentation system was a component system and thus nearly decomposable, that each step in the reaction was brought about by a specialized enzyme and did not depend upon other steps. Eventually researchers realized that the various reactions in the pathway were closely linked through a variety of connecting mechanisms; that is, they came to realize that fermentation relied on an integrated system. Thus, Chapter 7 first examines the development of mechanistic explanations that begin with direct localization, then looks at the advancement of a more complex mechanism that treated the system as a component system, and finally introduces an account of an integrated system.

In Chapter 8 we turn our attention to biochemical genetics. This field brought together a tradition of research on biochemical pathways with research on genetic control, focusing on the mechanisms by which genes expressed themselves in development. Mendelian genetics had developed a complex account of the genetic mechanisms that mediated heredity. This was a localizationist scheme. However, the model could not explain how genes expressed themselves in traits. This was recognized to involve a set of chemical reactions building up the ultimate products. The research program of biochemical genetics was directed at determining the pathways responsible for expressing the genes already identified through the Mendelian program. Ephrussi, Beadle, and Tatum investigated the biochemical processes involved in developing normal nutritional processes in *Drosophila* and *Neurospora*. The *Drosophila* work indicated the complexity of genetic control in development. It was still research based

on the Mendelian paradigm, attempting to discern genetic control from differences in the phenotypic expression. The *Neurospora* work provided information concerning the level of control. It began by inducing mutations and showing the phenotypic effects. It revealed, more specifically, that mutants lacked one or another component of the normal biochemical pathway of nutrition. Beadle and Tatum saw this as a new account of the complex mechanisms of genetic expression, on account which linked genes not with Mendelian traits, but with enzymes in biochemical pathways.

The result of Beadle and Tatum's analysis identifying genes with enzymes was that genes were no longer viewed as coding for Mendelian traits directly. Thus the localizationist story could not explain the expression of Mendelian traits, as it was originally intended. In this case, though, the consequence was not a rejection of localization and decomposition, but a *reconceptualization* of the phenomena to be explained. The research at the lower level forced a reconstitution of what researchers thought existed at the higher level; what seemed to be an appropriate account of the mechanism when all one had were the functional tools of Mendelian genetics was revised as a result of research directed at the lower-level analysis. We will see that specific lower-level constraints played a critical role in this transition.

CHAPTER 6

# Complex Localization

## 1. Introduction: Constraints on Localization

Mechanism was often seen as the only viable alternative to vitalism, and it was undertaken despite the fact that the simple mechanistic approaches were not wholly satisfactory as explanations, even on their own terms. At the same time vitalism can be seen as a consequence of mechanism, drawing sustenance from a clear vision of mechanism's limitations. Thus the mechanism of Gall, or of the preformationists, was mechanistic, but hardly provided a palliative for the explanatory needs it was designed to fulfill. As Flourens saw very clearly, simply relocating the faculties at a lower level was not an adequate explanation in the absence of an account of how the organs could do what they needed to. Simple, direct localization is by itself no solution, though it can serve as a preliminary stage bridging toward a more adequate account. When additional constraints are imposed, whether empirical or theoretical, they can serve simultaneously to vindicate the initial localization and to develop it into a full-blooded mechanistic explanation. In a somewhat less paradoxical form, additional constraints can result in a model that is more defensible, displacing and developing the initial localizationist picture. What we call complex localization is *localization multiply constrained*; that is, it proposes a set of components that contribute differentially to system function, and it incorporates independent constraints on allowable mechanisms from lower levels.

In addition to depending on multiple constraints, complex localization requires that we take decomposition seriously. Simple localization differentiates tasks performed by a system, localizing each in a structural or functional component. Complex localization requires a decomposition of systemic tasks into subtasks, localizing each of these in a distinct component. Showing how systemic functions are, or at least could be, a consequence of these subtasks is an important element in a fully mechanistic explanation. Confirming that the components realize those functions is also critical. Both are necessary for a sound mechanistic explanation.

As we shall see, the routes to complex localization are varied. Frequently a program of research begins with direct localization, which then develops into a more complex localization in which functional decomposition of tasks assumes a more central role. This is common in psychology,

where the legacy of Gall continues to influence theory development. Research often commences by dividing psychological activities into broad performance categories; for example, perception, memory, language, reasoning, and emotion. Noam Chomsky has provided one of the clearest expressions of this approach in his own "organology," strikingly reminiscent in tone of phrenology (cf. above, Chapter 4, section 2):

> We may usefully think of the language faculty, the number faculty, and others, as 'mental organs,' analogous to the heart or the visual system or the system of motor coordination and planning. . . . There seems little reason to insist that the brain is unique in the biological world in that it is unstructured and undifferentiated, developing on the basis of uniform principles of growth or learning—say those of some learning theory, or of some yet-to-be conceived general multipurpose learning strategy—that are common to all domains. (1980, p. 39)

Jerry Fodor has articulated the modularity, or organology, and generalized it beyond the domain of language, claiming that "modular cognitive systems are domain specific, innately specified, hardwired, autonomous, and not assembled" (1983, p. 37). These modular systems are concerned with a single domain and do not depend on the operation of other capacities in the process of executing their own specific functions. It is in roughly this sense that, for example, it has been held that syntactic processing does not depend on semantic interpretation, or that early visual processing is autonomous. Gall's phrenology also fits the profile as a theory postulating a modular structure to the mind. Destructiveness and combativeness are distinct faculties, after all.

Modularity seems inevitable, especially when dealing with a phenomenon as complex as mental life. If we do not divide up mental activities, or cannot discern their variety, we cannot even begin to see how they are realized. However, even when modularity is correct as a first approximation, it is only a beginning. It is imperative to know how the mechanisms work, and it is certainly not unreasonable to turn to the neurosciences to develop mechanistic models of psychological activities. It is in fact hard to see where else one might turn to gain a robust and realistic account of the mind. In this chapter we look at two examples from the neurosciences, one from the nineteenth and one from the twentieth century. Both are unrelentingly mechanistic.

It is important to recognize the eclectic character of neuroscience. There are variations in general strategy, in substantive methodology, as well as in domain; and all influence the character of the resulting models of cognitive functioning. Researchers vary, in particular, on the relative emphasis they place on cognitive models of performance as opposed to neurophysiological analyses. Some researchers are inclined to take cogni-

tive categories as relatively fixed and view the problem to be one of identifying neural structures responsible for the capacities those categories describe. For example, Norman Geschwind (1964, 1965, 1980) has elaborated a model for the neural substrate of language which emphasizes cognitive categories. He suggests that there are two primary structures implicated in speech: one concerned, roughly, with speech comprehension, and one with the control of speech output. Other researchers by contrast, are inclined to emphasize models based on information pertaining to neurophysiological processes. Thus Eric Kandel (Hawkins and Kandel 1984) and his collaborators have developed a model of learning in *Aplysia* which is motivated largely by detailed work on neurophysiological organization and, secondarily, by views about the mechanisms of associationistic learning. Likewise, Daniel Alkon (1989a, b; Farley and Alkon 1985) and his associates have a model of learning in *Hermissenda* which traces variations in learning to neural systems and associated cellular changes.[1] These differences are of emphasis and of degree, but they are differences that matter.

Even if the models depend on cognitive categories that are tentative and modifiable, they still significantly guide the search for neural mechanisms. The goal is, after all, understanding how the tasks *already identified* are performed. This does not deny that neurophysiological results might lead to the elaboration or modification of cognitive categories; they might even lead to wholesale revision or replacement, though we think this is certainly infrequent (cf. McCauley 1986; Wimsatt 1976). The point is, rather, that the models motivated by neuroscience cannot ignore cognitive psychology if the goal is understanding how cognitive tasks are performed; neither, despite philosophical imprimatur, can psychology ignore the neurosciences.

The solution is to seek robust, multiply constrained models. Psychological categories—or, better, *psychological organization*—provide a top-down constraint. Different conceptions of the proper decomposition of psychological functions impose different constraints on mechanistic models. Wernicke and Broca, for example, differed radically on which psychological organization was important. Broca embraced a faculty psychology, while Wernicke depended on an associationistic psychology. As we will see in the next section, an associationistic psychology leads to different mechanisms, or to different ways of conceiving of mechanisms, at the lower level. *Neural mechanisms* provide a bottom-up constraint, limiting the range of possible mechanisms that can explain psychological functions. Minimally this requires some consistency with what is known of neuroanatomy and neural pathways; more ambitiously it requires knowledge of neural mechanisms that have an appropriate internal organization.

Complex localization, at least in the cases we will describe in sections 2 and 3, results from an attempt to satisfy, simultaneously, constraints imposed at more than one level. This is what we mean by localization multiply constrained.

## 2. Top-Down Constraints

The differences, including the gains from multiple constraints, can be made clearer by returning to direct localization, and, in particular, to the natural heirs of Gall and direct localization. One of the more prominent French physicians to embrace the commitment to localized centers within the brain was Jean Baptiste Bouillaud (1796–1881), a student of Gall, founding member of the Société Phrénologique, and professor of clinical medicine at La Charité hospital in Paris. In 1825 Bouillaud described a patient who had lost the power of speech after a lesion in the anterior lobe, arguing that this was a direct confirmation of Gall's view that this lobe was the locus for articulate language. Bouillaud's method, very much like Gall's, relied on correlations; but, unlike Gall's, compared pathological symptoms with neurological lesions: Bouillaud regarded the examination of the dysfunctional brain as essential. His views met with considerable opposition, and after many public defenses Bouillaud was led to offer five-hundred francs to any person who could produce a case with no disturbance of speech in the presence of a lesion in the anterior lobes of the brain. No one collected on the bet.

This was the climate into which Paul Pierre Broca (1824–1880), as a defender of organology, entered the debate over localization. Although he rejected the cranioscopic method of phrenology, he did not reject the traditional emphasis on the size of the brain as indicative of intelligence, nor did he reject the reliance on cranial capacity in comparing races. He preferred, though, to measure brain capacity more directly, and he emphasized the pathological and clinical observation adopted by Bouillaud. Like Gall and Spurzheim, Broca was committed to direct localization. In sessions of the Société d' Anthropologie in 1861, and shortly thereafter at the Société Anatomique de Paris, Broca praised Gall and the principle of localization of function, describing it as "the point of departure for all the discoveries of our century on the physiology of the brain."

By August of 1861, Broca had described in considerable detail the anatomical changes accompanying a disorder of speech which he termed aphemia: a disturbance of "articulated speech" in the absence of any disturbance of the "general faculty of speech." The subject of the study came to be known as "Tan," following his habitual answer to questions. A congenital epileptic, Tan lost the ability to speak by the time he was thirty, but was otherwise healthy and intelligent. Over the ensuing years, how-

ever, Tan's case degenerated, with weakness and eventual paralysis of the right side of his body. Eventually, he died from gangrene. Through interviews with the hospital staff and other patients, Broca concluded that Tan's initial aphasia amounted only to a "loss of articulate language," with no comprehension deficit, no loss of hearing, and no loss of intelligence. Broca summarized the case as follows:

> From the anamnesis and from the state of the patient it was clear that he had a cerebral lesion that was progressive, had at the start and for the first ten years remained limited to a fairly well-circumscribed region, and during this first period had attacked neither the organs of motility nor of sensitivity; that after ten years the lesion had spread to one or more organs of motion, still leaving unaffected the organs of sensitivity; and that still more recently sensitivity had become dulled as well as vision, particularly vision of the left eye. Complete paralysis affected the two right limbs; moreover, the sensitivity of these two limbs was slightly less than normal. Therefore, the principal cerebral lesion should lie in the left hemisphere. (1861b, pp. 226–27)

An autopsy revealed widespread damage to the left frontal lobe. The tumor consumed the superior part of the frontal lobe, as well as much of the temporal lobe and other structures. As Broca described it, "If one puts back in imagination all that has been lost, one finds that at least three quarters of the cavity has been hollowed out at the expense of the frontal lobe" (ibid., p. 227). Assuming that the tumor spread uniformly, Broca concluded that its initial focus had to be in the frontal lobe, in particular in the third convolution of the frontal lobe (see Figure 6.1). The correlation of the symptoms and the projected course of the tumor led Broca to conclude that Tan's initial speech deficit was due to localized damage to the third frontal convolution and that, since other faculties were not initially disturbed, this must be a region specialized for language.

Embracing the "principle of localization" inherited from Gall, Broca boldly concluded from Tan's case that "the totality of the convolutions does not constitute a single organ, but many organs or many groups of organs, and there are in the cerebrum large discrete regions corresponding to large discrete mental functions" (ibid., p. 227). This is, of course, nothing less than Gall's organology. This famous case has earned Broca the reputation as the first to isolate a case of pure aphasia and to gain general acceptance in localizing a specific function in the brain (cf. Young 1970, ch. 4). His central contribution was probably a clear description of the phenomena associated with the lesions.[2]

It would not be much of an understatement to say that there was no novel clinical syndrome exhibited in the case described by Broca, and that the anatomical basis for his cortical localizations was meager indeed. Eventually, as a result of superior precision and technology, both meth-

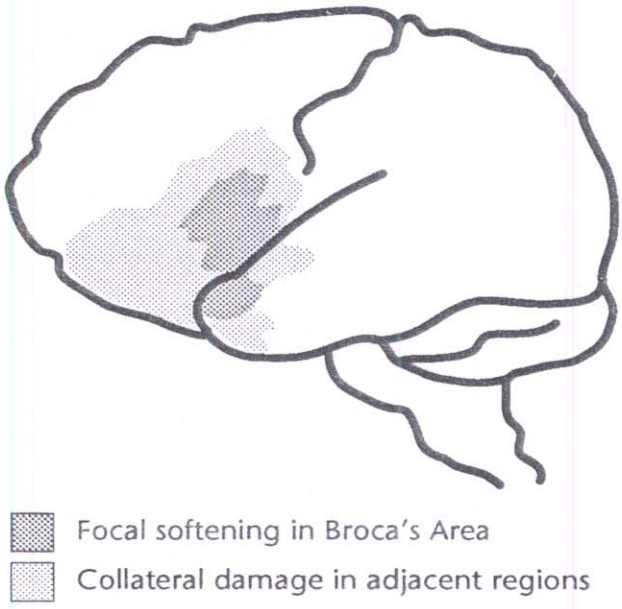



Figure 6.1. A Representation of the Damage to the Left Hemisphere in Tan, According to Broca (1861). Broca concluded, on the basis of the progress of Tan's aphasia and a post-mortem examination, that the damage in the focal area of the brain was responsible for Tan's condition and that that region was the "center for articulate language."

ods and answers improved. By examining Wernicke's views, and then more recent work on spatial memory, we should be able to see one path toward more complex localizationist models. This will occupy us in what follows.

### *Wernicke on Language*

Karl Wernicke's (1848–1905) work on sensory and motor aphasias is one of the clearest cases of research exhibiting the heuristics of decomposition and localization. Wernicke assumed that there was a series of discrete loci mediating the comprehension and production of speech; loci which, in concert, control behavior. There is, of course, interaction between these centers in the normal case: comprehension and expression are coordinated. The critical question from Wernicke's perspective, however, was whether it was *possible* for there to be normal comprehension in the absence of normal expression, or vice versa. He thought that the subsystems were independent despite their interaction, meaning that each could carry out its usual operations independently of the other; comprehension *could be* gained in the absence of expression, and expression effected in the absence of comprehension. Wernicke went on to assume that the subsystems also operated independently of one another in individuals not suffering from aphasia, that comprehension was actually independent of expression. The specific character of these *association centers*, and the functions they subserve, were interpolated on the basis of the specific clinical disabilities that resulted from injury to these centers. These behavioral deficits were viewed as the empirical cues to assessing psychological organization.

As we have pointed out, it is difficult to identify mechanisms when systems are operating smoothly, because components do not reveal themselves. The way the system breaks down often provides the best information available about the functional properties of the components. Thus, it is not merely that behavioral deficits are clues to *abnormal* functioning—they are clues to *normal* functioning as well. Functional independence, or decomposability, shows up if some capacities can be lost while others are unaffected. For example, loss of comprehension without a corresponding loss of expression, or vice versa, supports the claim to independence. On the other hand, there would be a failure of functional independence if there were *always* comprehension deficits pursuant to damage to the subsystem responsible for expression.[3]

These centers of comprehension and expression, however, were not the faculties posited by Gall and Broca. Wernicke assumed an associationistic model of learning and memory. Concepts are not independent entities, according to him, but the product of linking together more basic representations; the psychological unit becomes a set of associated ideas rather than an independent capacity. In Wernicke's own formulation, the possession of a "concept of a concrete object" is identified as a "constellation of memory images"; that is, it is a cross-modal association.[4] A concept of an observable object is thus identified with an association either between sensory representations, or between sensory and motor representations. Sensory impressions, or sensory representations, are modality-specific "ideas" or "representations" (*vorstellungen*), and they correspond to classical "proper sensibles." The concept of a rose, Wernicke says,

> Is composed of a 'tactile memory image'—'an image of touch'—in the central projection field of the somesthetic cortex. It is also composed of a visual memory image located in the visual projection field of the cortex. The continuous repetition of similar sensory impressions results in such a firm association between those different memory images that the mere stimulation of one sensory avenue by means of the object is adequate to call up the concept of the object. (1906, p. 237)

Given an associationistic slant, what are localized must be the centers of association rather than the cognitive faculties. A concept will not be localizable, since it will be no more than a constellation of associations; a faculty will not be localizable, since there are no faculties. The centers of association alone will possess discrete loci, and disruptions of these loci will in turn disrupt some aspects of behavior. In the case of language we must seek the associations built up in the process of learning to comprehend and produce speech; in speech perception, the simple ideas are acoustic representations; and in speech production, the ideas are the motor representations of speech (*Sprachbewegungsvorstellungen*)—that

is, they are the representations of the movements necessary for speech. Comprehension or expression will depend on the establishment of associations between simple ideas. These associations vary in complexity, from the primary reflex movements controlled by subcortical representations, to secondary motor movements initiated by cortically based representations, to voluntary control of behavior at the highest level (cf. Wernicke 1874, p. 95).[5]

Learning is, at least to a first approximation, understood as establishing an associationistic link between sensory impressions and motor representations. Accordingly, memory is a functional association between perceptual elements resulting from lowered resistance—or facilitation—following repeated, contiguous presentations. These include the sound/concept associations that are constitutive of sensory representations, the concept/word associations constitutive of motor representations, and the sound/word associations constitutive of inner speech.

The associationistic psychology embraced by Wernicke provided a psychological mechanism in terms of which aphasic syndromes could be interpreted. Any "higher" cognitive performance would be the result of an association of simple ideas. The conception of the psychological mechanism thus provided an account of what needed to be localized. The primitive ideas themselves, and especially the pathways that accomplished the association between them, were to be localized. To demonstrate that these were localized agents performing independent functions, Wernicke appealed to the different kinds of linguistic deficits (the aphasias) found in language comprehension and production. According to this account, specific aphasias were explained as breakdowns in an associationistic system; the diversity of the aphasias was a natural consequence of the complexity of this system.

Wernicke's approach allowed a systematic treatment of the various forms of aphasia within a single model, and that treatment provided the primary motivation for his model of language comprehension and production. It had been known for some time that localized destruction could impair comprehension of speech, while leaving both production and more general intellectual functions unimpaired. Wernicke noted a parallel phenomenon in which localized destruction could impair production of speech, while leaving both comprehension and general intellectual functions unimpaired. He saw that disruption of the first frontal gyrus would disrupt production; of the first temporal gyrus, comprehension (cf. 1874, p. 103). The former affliction has come to be known as *Broca's Aphasia*, and the latter as *Wernicke's Aphasia*.

Of course, in the vast majority of cases, what we see is an impairment in both categories as well as a variety of other cognitive deficiencies. Wernicke knew this well.[6] As he saw, however, the existence of complex defi-

ciencies is less important than the existence of the "pure aphasias"; that is, aphasias that affect strictly limited associations while leaving others intact. This is evidence that the representations can be dissociated, and therefore is evidence for decomposability within the system responsible for language. What most notably distinguishes Wernicke's program from Broca's, is this interpretation of the two primary forms of pure aphasia as sensory and motor deficiencies. As Wernicke understood sensory aphasia, there is a loss of the "acoustic memory representations," which provide the sound-concept associations. The meaning of heard speech therefore cannot be recovered. Articulate speech is intact; indeed it is copious and grammatical, though inappropriate to the context (cf. Wernicke 1874, p. 124). In motor aphasia there is a loss of the "motor memory representations" for speech, which provide the concept-word associations. Comprehension of speech is intact, but there is a loss of articulate speech (ibid., pp. 126, 130). In either case—for Wernicke, but not for Broca—the problem is not the loss of the concepts or the simple ideas, but an inability to establish the necessary associations.

As Wernicke saw it there are a variety of forms of pure aphasia, each with a characteristic syndrome and with very little overlap in symptomology. He wrote, even as early as 1874:

> It is self-evident that cases presenting two or three symptom-pictures in combination occur more frequently than the pure clinical forms caused by more or less isolated damage because the pathological process is generally very extensive. Undoubtedly, the typical pictures described, which adequately justify our formulation of a new clinical classification, actually do exist. . . . The mixed forms may readily be understood from the preceding discussion. (p. 114)

Clinical classification reflects the possible. The actual is more complex. Wernicke infers that there are centers responsible for these simple functions shown by the forms of aphasia, and these alone.

Wernicke's approach shows clearly how decomposition and localization depend upon one another. Information about the effects of localized disruptions shapes our conception of our basic psychological capacities. Clinical data is the motivation for distinguishing comprehension from expression, while circumventing the charge that the distinction between the two is but a false abstraction. On the other hand, views concerning psychological organization provide an account of what needs to be localized. Wernicke was limited to clinical data because he concentrated on human subjects in developing his model. There are other capacities that, within the last century, have become approachable with more sophisticated experimental techniques. It will be useful to look at a more recent case of localization based on an experimental rather than a clinical methodology. The general moral for our purposes is not substantially different.

*Spatial Memory*

In an interesting and impressive comprehensive study of hippocampal functioning, John O'Keefe and Lynn Nadel (1978) urge that the hippocampus is a structure specialized for spatial processing and memory. This is by no means the only proposal for understanding hippocampal functioning (for a general review, see Isaacson 1974), and O'Keefe and Nadel's proposal is far from unproblematic. In particular, David S. Olton and others have argued that the critical factor in hippocampal functioning is its involvement in working memory, rather than any involvement in spatial functioning (e.g., see Olton, Becker, and Handelmann 1979). It is not important to our project which, if either, of these proposals is correct, or which might finally be vindicated by further research. What is important is the methodological moral. O'Keefe and Nadel pursue a robustly characterized mechanism underlying spatial memory. They propose not only a localized capacity and an analysis of the relevant capacities, but a physical mechanism. As a result they propose a model that integrates information gleaned from more than one level and is subject to multiple, independent constraints.

O'Keefe and Nadel begin at a psychological level with the problem of explaining capacities for spatial orientation in humans and other animals. They distinguish two cognitive systems for guiding behavior; these systems provide alternative methods for solving spatial problems. What they call the *taxon system* provides an observer-relative, or egocentric, map of the environment. Activity guided by the taxon system is, accordingly, cued to external guideposts and bodily orientation. For example, an organism may elect to turn left at a given landmark in a maze when coming from a particular direction, just as we may be instructed to turn left at the second corner past the light. The *locale system* provides an observer-independent, or objective, map of the environment. Orientation is within an objective space. An animal may approach a position in the environment and, independent of the direction from which it approaches or any specific landmarks, still know its way around. For example, migratory birds can return to nesting grounds without depending on local cues, just as we may find our way to Chicago whether we begin in Minneapolis or Cincinnati.

An example used by O'Keefe and Nadel may help to introduce some operational rigor into the distinction. Tolman, Ritchie, and Kalish (1946) provided evidence for the deployment of objective cognitive maps by rats using a sunburst maze (see Figure 6.2). In the initial, training, phase (Fig. 6.2[a]), rats were taught to run a maze through a central chamber and along a side passage to a goal box with feeding stalls. After this phase the rats were released in a sunburst maze (Fig. 6.2[b]), which allows an array of directional choices. A *taxon* orientation would dictate that the rats se-

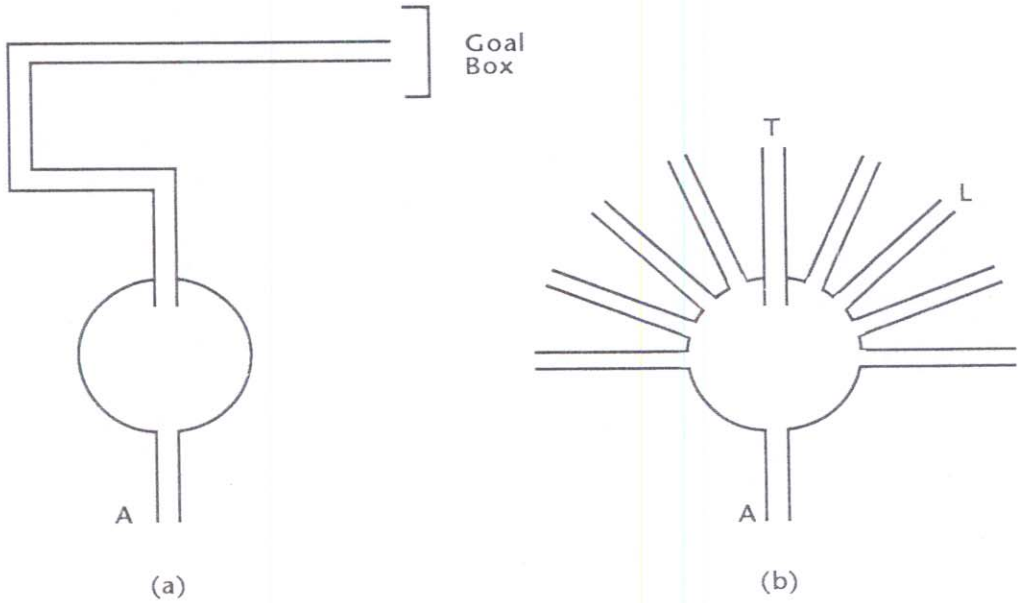



Figure 6.2. A Simplified Version of a Sunburst Maze Used by Tolman, Ritchie, and Kalish (1946). The rats were trained initially on the simple maze depicted in (a), taught to run through the central chamber from A, through the corridor, and to the goal box at the end. Subsequently, rats were released in a sunburst maze such as that illustrated in (b). Tolman, Ritchie, and Kalish reasoned that the path chosen in the second phase would reveal whether the rats had learned a specific route to follow or had learned the location of the goal. Rats uniformly took the direct route in the later trial.

lect the same passage (path T) they had been trained to use, as this would maintain their relative orientation and preserve response invariance. A *locale* orientation would require the selection of one of the paths providing a direct route toward the stimulus (ideally, path L), as this would reflect an awareness of location independent of immediate cues. In the experiment, the rats worked with a local orientation, leading Tolmon, Ritchie, and Kalish to conclude that rats had spatial maps.

O'Keefe and Nadel differentiate the two systems by turning to behavioral deficits associated with localized damage. They contrast two cases: "normal" animals (those which have undergone no surgical ablations) and "hippocampal" animals (from which the hippocampal structures have been bilaterally destroyed). O'Keefe and Nadel suggest that normal animals use both systems, while hippocampals use only a taxon system. This is based on the supposition that the locale system resides in the hippocampus, while the taxon system is localized in the parietal cortex.

The hippocampus, then, is the source of the *cognitive maps* that represent a nonegocentric, local-driven space (O'Keefe and Nadel 1978, p. 382). To make good on this claim, O'Keefe and Nadel must functionally dissociate the two, alternate, memory systems. In doing so they point to cases in which hippocampals perform less well than normal animals, em-

phasizing differences in the patterns of errors (1978, p. 382). In arguing for the claim that the hippocampus provides a cognitive, observer-independent map, it is crucial for the project that there be substantial decomposability. O'Keefe and Nadel explain their strategy in suggesting that "the hippocampus acts as a cognitive mapping system, which generates place hypotheses and exploration. Loss of this system forces the animal to rely on the remaining extrahippocampal system" (ibid., p. 90). The lesion studies that form the first, top-down leg of their analysis of hippocampal function depend on the assumption that the taxon system can function properly in the absence of the locale system (and presumably vice versa) and on a disparity between the performance of normal and hippocampal animals—that is, between normal animals and those with bilateral removal of the hippocampus and related structures. Assessing the difference between these animals depends critically on there being two functionally different cognitive systems controlling spatial orientation and on the system's discrete localization.

It is clear that, relative to normal rats, hippocampals are at a systematic disadvantage in running mazes. Although they uniformly perform worse than intact animals, it is *not* the higher frequency of gross errors that is interesting. It is not at all remarkable that, after extensive damage to their brains, animals with bilateral ablations do not function up to par. What is of special interest is the pattern of errors they manifest. For example, O'Keefe and Nadel cite a number of studies using mazes requiring spatial reversals; that is, "complex" mazes that demand, for completion, that the animal turn directly away from the goal (cf. O'Keefe and Nadel 1978, pp. 288ff.). Performance on mazes of this sort can be compared with performance on "simple" mazes which require no turns at more than right angles to the goal. A normal animal will be under the guidance of a locale system which contains an objective map of the environment. It will then orient toward the unseen goal. A normal rat's reliance on a locale system should make solution to the complex maze comparatively more difficult. It should follow that the rate of errors on complex mazes should be *greater* than on the simple mazes. But suppose, on the other hand, that a hippocampal animal lacks such an orientational capacity. It must therefore, following O'Keefe and Nadel, rely on taxon systems which are insensitive to objective orientation. We should then find no greater *relative* frequency of errors on the complex mazes, however great the gross number of errors may be. That is, hippocampals should perform worse than normals on both simple and complex mazes, but should show no significant difference between the two. And indeed, this is exactly what we find.

In one study by D. P. Kimble, the normals made three times as many errors on the complex as on the simple maze, even though the opportunity for errors was greater on the latter. Hippocampals, by contrast,

though significantly impaired on both mazes when compared to the controls, made twice as many errors on the simple mazes in comparison to the complex mazes. It would appear—or so O'Keefe and Nadel conclude—that hippocampal animals do not rely on the same hypotheses in problem solving as do the normal animals, and, lacking these hypotheses, they are less prone to error in cases where such hypotheses would be counterproductive; or, more carefully, the hippocampals' proportion of errors will not increase in circumstances that would make behavior based on the hypothesis prone to error.

The localization of the locale system in the hippocampus depends on the view that the functions of memory are decomposable and can be dissociated so that, in some cases, they function independently. O'Keefe and Nadel are explicit on the point:

> It appears that there are different types of memory, relating perhaps to different kinds of information, and that these are localized in many, possibly most, neural systems. . . . [There] is no such thing as the memory area. Rather, there are memory areas, each responsible for a different form of information storage. (1978, pp. 373–74)

Complexity once again is treated as multiplicitly.

O'Keefe and Nadel thus assume there is decomposability at more than one level. First, there must be substantial independence of systems involved in memory from other, possibly related, cognitive functions. If damage to the locale system impaired general cognitive functioning, we would expect to see a more complex mosaic of symptoms than hippocampal damage alone would suggest. Second, there must be substantial independence of the various components of memory from one another. The taxon system is supposed to execute its characteristic functions even in the absence of structures that realize the locale system.

Because O'Keefe and Nadel claim that different neural mechanisms are involved in these two spatial systems, it is critical that the two systems are both functionally independent and localizable in different neural structures. This is straightforwardly inconsistent with equipotentiality of neural structures. Equipotentiality would allow alternative regions of the brain to assume functions they do not normally realize and so would undermine the specialized character of the posited mechanisms. For example, young children (from roughly 20 to 36 months) who suffer traumatic aphasia, effectively recover language skills. "In the very young," Eric Lenneberg tells us, "the primary process in recovery is acquisition, whereas the process of symptom-reduction is not in evidence" (1967, p. 150). If there were a potential for recovery of function in this same sense for memory as well as language, then any appeal to the neural apparatus specific to the hippocampus would be pointless, for there would be no

useful way in which to have neural mechanisms specialized for particular functions. This is exactly the conclusion O'Keefe and Nadel recognize, and they contest an interpretation resting on equipotentiality as the basis for recovery of function:

> The take-over notion would appear unlikely on purely anatomical grounds. Although one bit of cortex might conceivably substitute for another, the unique machinery of a structure such as the hippocampus is not so easily replaced or simulated. . . . [Most] tasks put to experimental animals can be solved through the use of any of several different strategies, some of which would be dependent upon different brain structures. Recovery of function could thus represent a switch to alternative modes of solution dependent upon intact brain tissue, rather than the actual reorganization of brain tissue. (1978, p. 235)

The denial of recovery of function in favor of a "switch to alternative modes of solution" does much to reinforce the claim to specialization and therefore to localization.[7] To the extent that we see functionally specialized neural systems, we should expect to see localized functions.

## 3. Bottom-Up Constraints

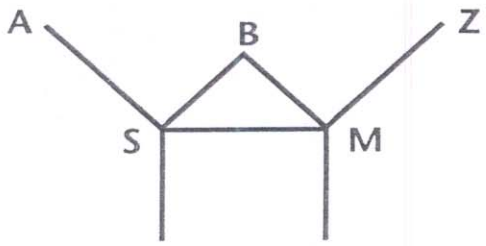
The cases discussed in section 2 focus on the analysis of behavioral deficits and the correlation of these with physiological destruction. They are, in fact, reminiscent of the cases discussed in Chapter 4. Nonetheless, they do exhibit marked advantages over the simple localization of Gall, insofar as they introduce more direct behavioral and physiological evidence as well as use pathological data. Their goal is to discern neural correlates of psychological activity. Still, the evidence appealed to in section 2 is still simply correlational: studies have the goal of finding the various mental states and/or behaviors correlated with specific forms of neural functioning. Behavioral deficits are correlated with neural damage. Broca's analysis of "Tan" is, effectively, limited to this stage. There is no independent evidence about the underlying neural structures or their capacities. As we have said repeatedly, this kind of evidence is not sufficiently probative to underwrite any strong conclusions about the character of the neural mechanisms, even though it may be enough to launch a program of research. Correlational studies finally need to be coupled with analytical studies geared to determining how mechanisms consistent with known anatomical and physiological features produce the phenomena we see. It is only once the research program is coupled with these analytical studies that it promises to *explain* the observed correlations and psychological/behavioral capacities; or, at a minimum, to show that it is reasonable to believe there are discrete neural mechanisms responsible for such correlations. Let us see how the cases discussed in section 2 were extended by the introduction of lower-level constraints.

*Wernicke on Language*

Wernicke settled for the less ambitious goal of showing that it was reasonable to believe there are appropriate neural mechanisms. Arguing for the consistency of the correlational studies with known neuroanatomical facts, he hoped to make it plausible that a mechanism could be found. As we emphasized in section 2.1, Wernicke's psychological decomposition was grounded on an associationistic model of learning and memory, which essentially treated the memory trace as an association, partially defined in terms of sensory-motor associations. Wernicke localized the motor-control region for speech in the anterior lobes (specifically, the first frontal gyrus), and the sensory region in posterior lobes (specifically, the first temporal gyrus). This allowed him to analyze the aphasias as the result of disruptions in normal sensory-motor control.

His choice of loci was in no small part supported by Meynert's physiological taxonomy. Bell and Magendie had shown that, within spinal nerve roots, there was a "specific action" assigned to individual nerve pathways: anterior nerves were specialized for motor functions, and posterior nerves for sensory functions. Meynert generalized the view: Assuming that the function of the cortex was to effect sensory-motor coordination, he projected the basic dichotomy observed by Bell and Magendie to the division between posterior and anterior cortical structures. Meynert distinguished three pathways: the "association pathways" which "begin and end in the cortex" and are responsible for intracortical communication; and two projection pathways which "begin and do not end in the cortex" and are, respectively, responsible for input to and output from the cortical structures. He concluded that sensory functions were to be assigned to the posterior, centripetal, pathways, while motor functions were to be assigned to the anterior, centrifugal, pathways.[8] Wernicke continued to carry the *connectionism* and *associationism*, with their physiological interpretations, into the analysis of the higher functions. The aphasic syndromes would then arise from the disruption of the associations, now conceived as the association pathways within the nervous system.

In the earlier work of 1874, Wernicke recognized three forms of pure aphasias[9]; that is, he recognized three forms of aphasia that exhibited the symptomologies of the pure type. These can be seen by turning to Figure 6.3. A *pure sensory aphasia* (or Wernicke's Aphasia) resulted from the loss of the association between *acoustic representations* of speech (S) and the concept (A), leaving the subject able to talk but unable to comprehend speech. A *pure motor aphasia* (or Broca's Aphasia) resulted from a dissociation of the *motor representations* (M) from ideation (Z), leaving comprehension intact but expression impaired. Finally, a *pure conduction aphasia* would result from a dissociation of the acoustic representations (S) from motor representations (M), leaving both comprehension and expression




(S) Acoustic Imagery
(M) Motor Imagery
(B) Concept Center
(A) Discharge (Initiation of Conceptualization)
(Z) Ideation (Goal Planning and Motor Integration)

Figure 6.3. Wernicke's Scheme for Interpreting the Various Aphasias. Wernicke interpreted the various forms of aphasia in terms of the disruption of specific association tracts. Disrupting the connection between S and A caused a comprehension deficit; disrupting the connection between Z and M produced an expression deficit; and disrupting the connection between S and M resulted in a conduction aphasia. In 1874 these were the only "pure aphasias" recognized by Wernicke, with various mixed aphasias resulting from simultaneous disruption of more than one association tract. Wernicke subsequently recognized more forms of aphasia.

intact, but leaving us unable to mimic heard speech. In terms of Meynert's physiology, the pure conduction aphasias would be the result of disrupting association pathways, while sensory and motor aphasias would be the result of disrupting centripetal and centrifugal pathways.

The general taxonomy of the aphasias was thus at least consistent with the extant neuroanatomical information. This was more than a simple, direct localization, since it proposed a functional decomposition and showed this to be consistent with neuroanatomy and physiology. It remained to be proved that the neural centers were indeed capable of performing the functions assigned to them.

### *Spatial Memory*

O'Keefe and Nadel adopt a more ambitious strategy in supporting localization, arguing that both the functional properties and the anatomical structure of the hippocampus, as well as its place in more general neural structures, are consistent with the hypothesized functions of the hippocampus. They thus argue not only that there is reason to think there are appropriate neural mechanisms, but that the detailed structure of the hippocampus promises to realize the needed functions. Here we find at least two discrete substages in O'Keefe and Nadel's argument: a correlation of psychological activity with detailed neurophysiological functioning, and an explanation of this correlation in terms of the structure of the hippocampus. The first of these substages is still correlational, though carried on at a level of greater detail; the second is no longer merely correlational, but overtly mechanistic.

In the first of these substages, one target phenomenon is hippocampal theta.[10] Theta is a rhythmical, slow activity characteristic of EEG during, for example, exploration, or in the presence of novel stimuli. It serves as

a measure of the relative involvement of neural structures. In Figure 6.4, theta activity is especially evident during swimming and the indicated periods of movement. In a summary form, O'Keefe and Nadel (1978) claim that

1. Hippocampal theta is correlated with activity in systems implicated in movement. It is not merely a matter of general arousal or increased motivation (cf. p. 174).
2. Theta is not the result of motor feedback. Peripheral control is ruled out because theta activity can be enhanced even under paralysis (p. 175).
3. Theta must be correlated with some "more global or molar aspect of behavior" (p. 176); it is not limited to the activation of any one specific muscle or muscle group.
4. High-frequency theta is correlated with speed and distance traversed. Looking again at Figure 6.4, the frequency differences in the two periods of movement are correlated directly with these variables (p. 182).
5. Low-frequency theta appears to be correlated with attention and exploration, rather than movement (p. 185); thus, theta presents a complex picture of neural activity *somehow* correlated with movement, but not with *any specific* motor activity.

One of the most salient correlates of hippocampal activity is location. Demonstrating this requires descending to a lower level of analysis, as EEG only reveals more macroscopic phenomena. O'Keefe and Nadel turn

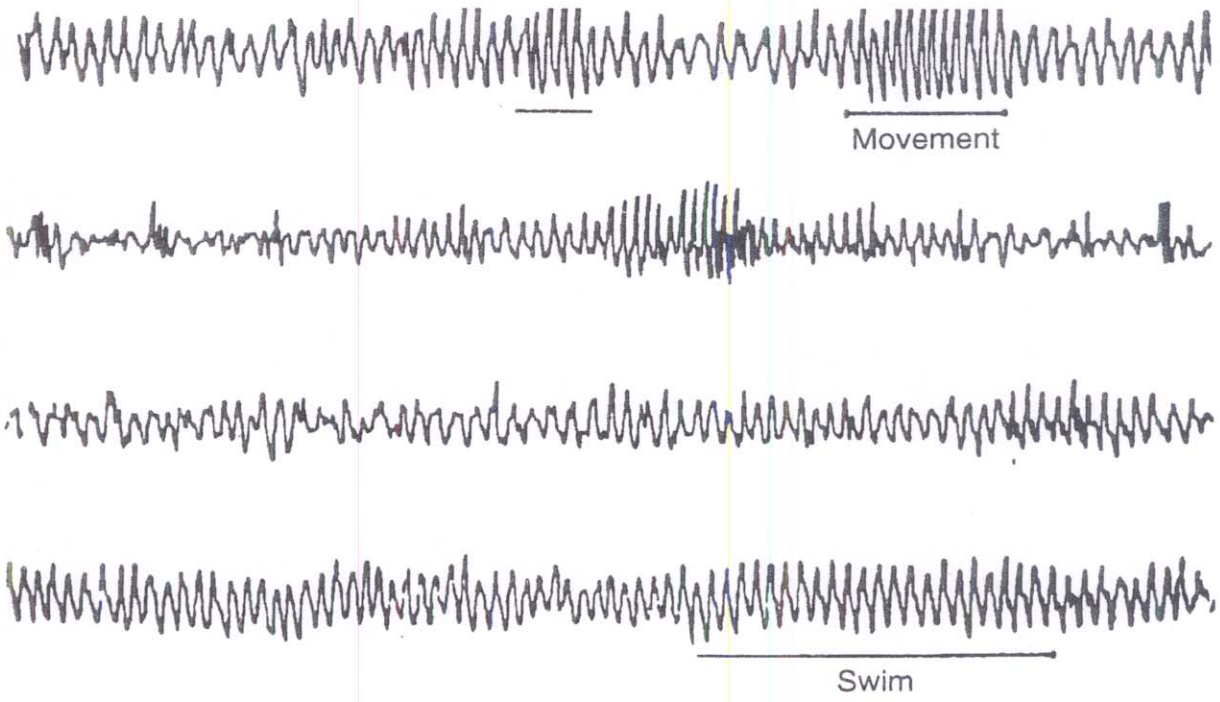



Figure 6.4. A Simplified Illustration of Theta Activity from EEG Readings. The rhythmic, slow activity characteristic of theta is especially evident in the periods of movement and swimming.

to single-unit recordings in the hippocampus. These display a striking correlation between location in a test apparatus and spiking in the appropriate cell. This spiking is limited to specific locations and independent of the type of activity. For example, as a rat reaches a choice-point in a maze, there will be correlated activity in particular cells that is independent of such factors as the direction from which the point is approached. Such correlations are, at least, consistent with the claim that the hippocampus is specialized for some form of spatial memory.[11]

The EEG correlations, and the correlation of hippocampal activity with features that would be required of a local system, are important, but the explanation of these correlations in terms of the structure of the hippocampus is critical to show that we are not engaging in a "vacuous functional localization" (Bechtel 1982)—if we are to show, that is, that the hippocampal structures are the *actual* mechanisms underlying behavior rather than merely potential mechanisms. This takes us to the second substage, emphasizing bottom-up constraints.

Neuroanatomically, the hippocampus has two principal components: the *fascia dentata* and the hippocampus proper (the complete hippocampus is formed by these two interlocking structures, as illustrated in Figure 6.5). Both of these structures are relatively simple. At a cellular level each is fundamentally a single sheet of large neurons with projections in parallel. In the fascia dentata, the primary neurons are granule cells; in the hippocampal regions—CA1, CA2, and CA3—they are pyramidal cells. Within the hippocampus proper (and to a somewhat lesser extent, the fascia dentata), basket cells serve locally as interneurons, evidently subserving an inhibitory effect. The conduction of impulses is, to at least a first approximation, unidirectional, with impulses flowing from the fascia dentata through CA3 to CA2 and, finally, to CA1. The most important afferent tracts (see Figure 6.6) to the hippocampus come from the entorhinal cortex (EC), which exerts a strong excitatory effect on the granule cells of the fascia dentata, and from the medial septum (MS), which projects to cells throughout the hippocampus and the fascia dentata.[12]

Given this structure, O'Keefe and Nadel advance a model of a locale system that is both elegant and straightforward. There are five primary elements, three of which are localized within the hippocampus proper. Inputs to the hippocampus come from the entorhinal cortex and the medial septum. The former controls sensory inputs or configurations of sensory inputs; the latter modulates theta frequency in response to motor control, including an animal's speed or the distance covered. The *initial coding* takes place in the fascia dentata. Units in this region respond to specific sensory inputs, or to configurations of such inputs, and are "sensitive not just to the simultaneous presence of two or more stimuli, but to their occurrence in a unique spatial configuration" (O'Keefe and Nadel

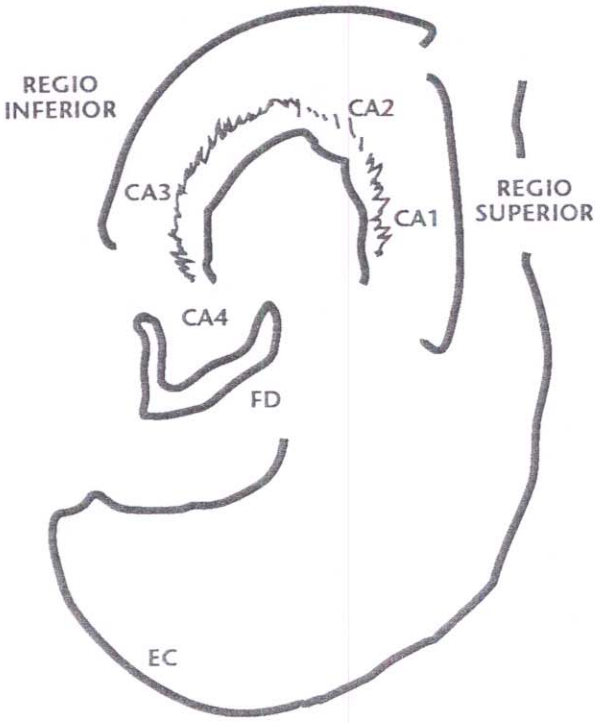



Figure 6.5. A Diagrammatic Representation of Hippocampal Structures. The hippocampus is a region of cortex consisting of two interlocking structures, the fascia dentata (FD) and the hippocampus proper; the latter can be differentiated into three further regions (CA1, CA2, and CA3). It is generally accepted that the hippocampus is critical for encoding of memories, and O'Keefe and Nadel speculate that it is specifically involved in spatial memory.

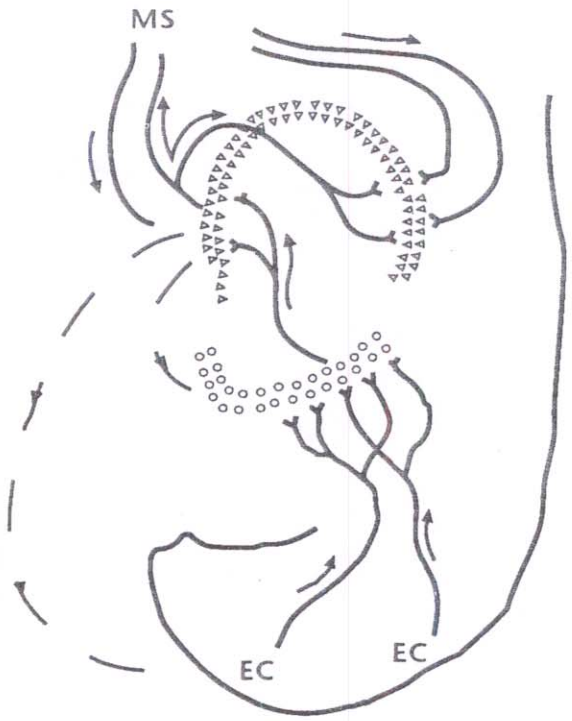



Figure 6.6. A Second Representation of Hippocampal Projections. This is an extremely simplified representation of projections involving the hippocampus and related structures. The actual connections are far more complex and greater in number than the figure suggests. Inputs come from the entonrinal cortex (EC) and the medial septum (MS), as well as a variety of other extra-hippocampal cortical regions.

1978, p. 221). The output of the fascia dentata is to a *place system* localized in CA2 and CA3. The function of this system is to use the input from the fascia dentata to determine and represent the current place of the animal, and then to represent this in such a way that the current place is located within a system of places having a determinate distance and direction from the current location (cf. O'Keefe and Nadel 1978, p. 223). The place sys-

tem must tell us where we are, *relative to* other (known) places. For the construction of such a "map," theta rhythms are thought to be crucial. O'Keefe and Nadel conclude (1978, pp. 157ff.) that the medial septum serves as a "pacemaker" designed to impose a synchronous bursting pattern on the hippocampus and the fascia dentata. Within the fascia dentata, this means there will be a pulsed output to CA3 capable of activating only certain pyramidal cells; for example, one can envision a pulse traversing a series of pyramidal cells, activating only the set it reaches at the time that inhibitory processes are minimal. Since theta frequency varies with the rate of spatial displacement, which cells within CA3 are activated will depend, in part, on how rapidly the animal is moving from place to place. A pulse responsive to a particular configuration of stimuli, though projecting to a series of pyramidal cells, will activate only a narrow band. The band activated will then correlate place with output, and relative location on the series will represent external place relative to other locations.

The final component of O'Keefe and Nadel's model is a *misplace system*. It is located, somewhat tentatively, in CA1. Its task is "to generate mismatch signals when the representations of places do not match the stimuli experienced in the corresponding part of the environment" (1978, p. 228). The misplace system consists of units that have higher rates of firing when there is a mismatch: the higher the rate, the more extensive the mismatch. It is not clear to us, in any detail how this mechanism would work.

Appealing to lower levels allows O'Keefe and Nadel to offer an admittedly partial and provisional model of performance based on the organization and function of the hippocampus. The model is no longer merely correlational, but the product of both a higher-level decomposition and a localization constrained by neurophysiological information. In this case, the behavioral features focus on the deficits of hippocampals relative to normals, and the specific pattern of errors of both groups. The resulting functional decomposition is admittedly speculative, but it plays an important role in the program of research. At a finer level of detail, integrating both physiology and behavior, we have correlational information about the relation of behavioral variables to large- and small-scale neural activity. And at a yet finer level we have neuroanatomical and neurophysiological information about the components of the hippocampus, providing evidence that the hippocampus has the computational resources to perform the function assigned to it. While every element of the case, as often is true, is indirect and inferential, that supports no objection whatsoever. What is important for our concerns is that O'Keefe and Nadel supplement the initial behavioral information with evidence garnered at lower levels. The result is a picture of the hippocampus as a complex machine with different parts contributing in characteristic and distinct ways to the overall task. This is localization under multiple independent constraints.

## 4. Conclusion: The Rise and Decline of Decomposability

The cases described in this chapter can justly claim to have circumvented the central charges against direct localization. As we have seen, opponents of Gall, such as Flourens, were quick to press that his own explanations of behavior were spurious, as they explained behavior in exactly the terms they described it; such explanations were no more than redescriptions. Moreover, they were metaphysical and speculative if they were meant as postulations of occult powers. One central feature of complex localization is that it utilizes a complex strategy in evaluating and developing models; it simultaneously relies on evidence concerning phenomena pitched at different levels of organization. The cases in this chapter provide evidence relevant to psychological processes, and *independent* evidence relevant to neurological functioning. As a consequence, the resulting model is designed to synthesize and accommodate phenomena at more than one level; it acknowledges constraints from two independent directions. Thus, Wernicke's model of language processing depended on a detailed characterization of the aphasic symptomology, because that provided evidence concerning patterns of association. It depended as well on information from Meynert's neurophysiology. Similarly, O'Keefe and Nadel's characterization of the hippocampus as a system for spatial representation relies on behavioral data designed to support a characterization of our capacities and a claim to their independence. It also depends on evidence regarding neurophysiological functioning.[13]

A second feature of complex localization is that, at some point, the researcher does not simply divide up the tasks performed by a system, but rather divides a given task into subtasks that interact in the performance of the overall task. As we will explain in substantially more detail in Chapters 7 and 8, the most natural assumption is that the interactions are linear: one subsystem performs a specific task, and that product serves as the input to the operation at the next stage. In Wernicke's model this decomposition is found in the differentiation of comprehension, production, and association; in O'Keefe and Nadel's work it is found in the differentiation of distinct tasks within the functional organization of the hippocampus. No one task is a primitive or fundamental activity in the sense the phrenologists proposed; each is a complex and composite activity achieved by distinct subsystems realizing their own intrinsic functions. Decomposition of the task, with further decomposition into subtasks in light of localization, is what finally yields a mechanistic explanation.

In the cases discussed in this chapter, the functional decomposition and localization, as it evolved under multiple constraints, maintained much of what a simple localization would have projected. If an initial decomposi-

tion and localization do lead to a recognition of one sole, or centrally, responsible component, then, at least as a first approximation, that component can be treated independently. This is what we call *first order independence*. It is one natural outcome of a program of decomposition and localization, and it sets a clear agenda for a research program. We represent this process in Figure 6.7. Figure 6.7 also represents a choice-point in developing a research program: whether to go to a lower level of organization in elaborating the mechanism. When one does integrate a lower level, subcomponents within the system need to be identified and isolated, and their functional properties ascertained. The nature of the interactions of these subcomponents with other systems and components of other systems will evidently also need to be examined, thereby partially compromising the assumption of independence. This is why we refer to the independence as *first order*.

Carrying through this program of research will require operating at more than one level of organization. O'Keefe and Nadel incorporate psychological data, information concerning larger neural structures, as well as a variety of other information, including information about the functional properties of individual cells.

It is in the construction of theories that bridge levels that the importance of decomposition and localization is most clearly manifested. The synthesis of information from more than one level relieves us from methodological limitations of more restricted studies. Though it is often hard enough to find one model adequately capturing the psychological and behavioral data if we are restricted to this data, it is even more difficult to know whether we are confronted with statistical or experimental artifacts. From the other direction, information about neurological functioning in the absence of an interpretive framework can be simply overwhelming. To highlight the point it may help to recall the interpretive principle that brought lesion studies within the compass of O'Keefe and Nadel's purview:

> In the absence of the locale system, hippocampal animals solve problems with taxon hypotheses, and their behavior reflects the properties of the taxon systems. Certain experimental situations will primarily reflect the operations of the taxon systems. . . . In these and other taxon-based situations we would expect the hippocampal animal to perform at least as well as the intact animal. (1978, p. 316)

In their subsequent treatment of operant and classical conditioning, O'Keefe and Nadel explain the lack of a clear *general*—that is, a nonspecific—deficit in hippocampal animals by pressing that the "solution" is available using only taxon hypotheses and according the locale system a "minor" but unspecified role. Many problems, then, can be solved in dif-

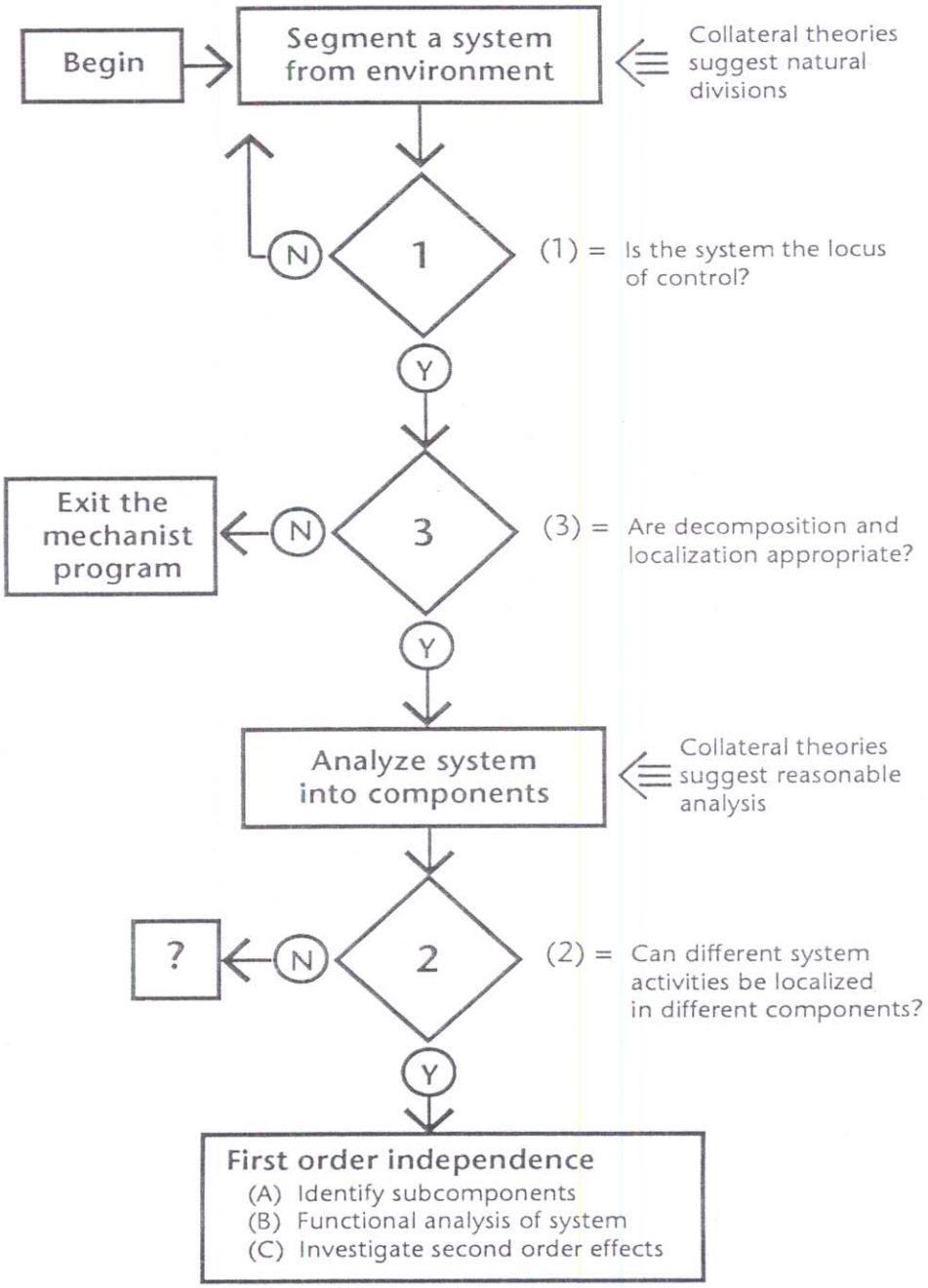



Figure 6.7. One Result of Localization and Decomposition. In some cases a result exhibits first order independence; that is, the primary dimensions involved in system behavior can be explained in terms of the behavior of a single component, in light of a functional analysis of the system. The resulting program of research would extend the analysis by elaborating the mechanisms and investigating the interaction with other components within the system.

ferent ways. In experimental settings we can hope to pull apart the systems and isolate their behavior.[14] This is possible only if the system is either decomposable or nearly decomposable and only by imposing complex constraints.

The critical question is whether there is a decomposition of cognitive functioning that respects neural localization. If there is a failure of localization and decomposition, we can make adjustments at either level, or at both. We could, for example, reject the higher-level, functional characterizations in the face of a failure of localization; that is, failure of localization can be treated as evidence against the higher-level functional analysis. The problem, thus understood, lies in the (potential or actual) mismatch between function and structure. On the other hand, it may turn out that the simple aggregative or linear structure congenial to decomposability and exemplified in the phrenological tradition, as well as the views of Wieland and Warburg on cellular respiration, is untenable; that is, the functional decomposition may fail. These possibilities will occupy us in the next two chapters.