Strategies for Analyzing Complex Organization in the Nervous System: I. Lesion Experiments

Paul Grobstein

In a 1971 paper entitled "Nerve Cells and Behavior," Donald Kennedy provided a picturesque and still instructive metaphor for the state of neurobiology as he perceived it at the time. Kennedy invited readers to consider the Astrodome during the opening game of the 1971 World Series, together with an expedition of interplanetary visitors whose task it is "to figure out the basic program for the activity taking place inside this noisy hemisphere—or, in other words, to learn the rules of the game." Two competing technologies evolve: one is based on highly sensitive parabolic receivers that, positioned on the surface of the dome, record sound from large areas of the stadium at one time, and the other relies on tiny, short-range microphones insinuated inside where they monitor individual sound sources. "An odd feature of the spontaneous division between the groups of investigators is that each is agressively chauvinistic about its own technology. Members of the second team find the technical approach of the others crude and unselective as compared with the more refined activity of microprobing. Those of the first group regard microprobing as an appropriate mechanical sublimation for people who can't handle the mathematics of computers. . . ." "This is perhaps not an entirely grotesque caricature of large-systems neurophysiology," Kennedy went on. "The electrical analysis of units en masse always obscures the borderlines between functional groupings, and gives a muddy picture of what is going on. On the other hand, attempts to synthesize from single-element data confront one with the dilemma that much of the recorded activity is irrelevant to the behavior studied, and that functional populations are thus difficult to construct from individual elements."

Kennedy developed his metaphor in defense of an "embattled minority" committed to the analysis of invertebrate nervous systems, where it was presumed that the existence of smaller numbers of reproducibly identifiable neurons would facilitate the identification of functional populations based on the properties of individual elements. Then a youngster in that minority family, I could not but be impressed by Kennedy's eloquence, but he was certainly neither the first nor the last to argue that brain function and behavior would be clarified by analysis of individual neurons and their interconnections. The proposition was strongly defended by Roger Sperry in the 1940s, against the countervailing, and then dominant, more global approaches of a number of investigators, including Paul Weiss and Karl Lashley (see Grobstein 1988a). With the advent of the microelectrode, and modern neuroanatomical tracing techniques, the promise of a close isomorphism between behavioral and neural properties began to seem real (cf. Barlow 1972). Indeed, by the mid-1970s, invertebrates on longer needed a defense, and the "simple systems" approach which they had come to epitomize was the dominant force in analysis of the vertebrate nervous system as well.

"But somewhere underneath, something was going wrong. The initial discoveries of the 1950s and 1960s were not being followed by equally dramatic discoveries in the 1970s.... None of the new studies succeeded in elucidating the *function* of the visual cortex." (Marr 1982; italics his) David Marr was not the only one to notice a problem. Discomfort, if not always openly acknowledged, has been increasingly expressed by neurobiologists of almost every sort, including those working on simple systems (Davis 1976; Selverston 1980; Mpitsos and Cohan 1986; Loeb 1987; Eaton and DiDomenico 1985). While diagnoses of the problem vary in detail, a common theme is that characterization of functional populations is not only "difficult to construct from individual elements" but frequently impossible, even in many simple systems. The difficulty has turned out not to be solely neuronal number but something deeper and more akin to Kennedy's prescient concern about determining whether observations made are relevant to the behavior being studied. So long as neurons and neuronal networks seemed to be displaying properties clearly isomorphic with behavior, it was possible to ignore the logically prior issues of how one knew what to look for in the behavior of neurons, and of how one established the relevance of particular neuronal properties for behavior. Unfortunately, or fortunately if one has a taste for the unknown, what has emerged is an unanticipated degree of neuronal complexity. There are more neuronal networks and neuronal properties than one seems to need to account for central pattern generation, more maps than one seems to need to account for vision, more circuits than one seems to nee
to account for escape behavior, and so forth and so on.

What the last 10 or 15 years of research on the nervou
system imply is that neurobiologists have tended t
underestimate not only neuronal complexity but b
havioral complexity as well. Stereopsis, to cite but on
example, is not simply a problem of converging inpu
from the two eyes; it requires as well a determination
which inputs to converge. This and a variety of simil
examples have motivated a healthy new look at th
computational problems involved in particular acts
behavior (cf. Arbib 1975) and in so doing have provid
some explanation for the degree of complexity observ
in the nervous system. Marr, an early and forcel
proponent of this perspective, made explicit the approac
distinguishing between three levels of analysis of
information processing device: computational theor
representation and algorithm, and hardware implement
tion (Marr 1982). Given that Marr too was writing at t
time as the advocate of a minority position, it is r
surprising that he gave computational theory pride
place among the three levels of analysis.

> *Each of the three levels of description will have its place in*
> *eventual understanding of perceptual processing, and of cou*
> *they are logically and causally related. But an important po*
> *to note is that since the three levels are only rather loos*
> *related, some phenomena may be explained at only one or t*
> *of them. ... It is the top level, the level of computational thec*
> *which is critically important from an imformation-process*
> *view. The reason for this is that the nature of the computati*
> *that underlie perception depends more upon the computatio*
> *problems that have to be solved than upon the partic*
> *hardware in which their solutions are implemented. To phr*
> *the matter another way, an algorithm is likely to be underst*
> *more readily by understanding the nature of the problem tc*
> *solved than by examining the mechanism (and the hardwa*
> *in which it is embodied.... Trying to understand perception*
> *studing only neurons is like trying to understand bird flight*
> *studying only feathers: it just cannot be done.*

An uncritical reader could be forgiven for believ
that Marr advocated throwing out the baby with the b
water, forgetting the nervous system entirely until
computational and theoretical groundwork necessary
making sense of it has been laid. "Neuroanatomy,
example, is clearly tied principally to the third level,

physical realization of the computation. Neurophysiology, too, is related mostly to this level ... one has to exercise extreme caution in making inferences from neurophysiological findings until one has a clear idea about what information needs to be represented and what processes need to be implemented." Significantly, however, Marr's own seminal analysis of visual processing as a computational problem was, as he himself acknowledged, "very much influenced by the fascinating accounts of clinical neurology.... Particularly important was a lecture that Elizabeth Warrington gave at MIT in October, 1973.... Warrington's talk suggested two things. First, the representation of the shape of an object is stored in a different place and is therefore a quite different kind of thing from the representation of its use and purpose. And second, vision alone can deliver an internal description of the shape of a viewed object, even when the object was not recognized in the conventional sense of understanding its use and purpose."

What is noteworthy in Marr's recollections is not only the origin of a significant computational analysis in observations of the nervous system, but the particular character of those observations and of the kinds of conclusions that can be drawn from them. Marr recognized in studies of behavior following brain lesions a capability to make meaningful statements about the computational problems the nervous systems is designed to confront as well as about the strategies it employs. This alone would justify some consideration of lesion studies in a discussion of computational neuroscience. In fact, lesion studies have played and, I will argue in this chapter, will continue to play an equally significant second role: they represent a natural way to deal with complex populations of neurons, and exemplify a needed intermediate level approach (Arbib 1985; Grobstein 1987, 1988b,c, and chapter 19 below) to problems of information processing in the nervous system. Properly conceived and interpreted, lesion experiments provide a basis for characterizing "functional populations" of neurons and exploring the interactions among them. The behavioral deficits resulting from localized damage not only speak to the relevance of particular neuronal populations for particular behaviors but also make it possible to identify distinguishable "information processing blocks," of use in describing both nervous system organization and behavior. The level of analysis at which meaningful isomorphisms between neuronal organization and behavior exist emerges from the observations, rather than having to be presumed at the outset, as is necessary for both computational and neuronal characterizations. What looks inordinately complex and indeterminate from the single cell level is frequently more orderly and approachable in terms of the intermediate level concepts made available by lesion studies.

Lesion Studies: The Need for a Defense and a Logic

The rapid expansion of neuroscience in the last two decades has led to important advances in research methodology that antiquate the lesion techique.... Thus, because the macrotechniques of neuroscience research have become disused in favor of the microtechnology of the 1970's and 1980's, in recent years the study of the effects of brain lesions on behavior has come under widespread criticism.... There is a problem in neuropsychology, however: it is not feasible to use most of the modern microtechnology when studying human subjects. Furthermore ... the most direct technique for studying the effects of brain lesions in humans is still to study the effects of analogous lesion in nonhuman animals. Thus the lesion technique continues to be an important tool.... But its use is fraught with considerable problems. (Kolb and Whishaw 1980)

Many, perhaps most, modern neurobiologists would characterize lesion studies in much the same terms as used by Kolb and Wilshaw: the methodology is antiquated, has severe interpretational problems, is a form of inquiry suitable at best only for generating hypotheses, and, even then, is excusable only in cases where more recently developed techiques cannot be applied. In this context, it would be unreasonable for me to expect most readers to take seriously the claims made in the introductory section for the value of lesion experiments without providing some explanation for the difference between such claims and the current disrepute into which the lesion methodology has fallen. Here, I must confess that, like Kennedy and Marr, my concern for defending a minority position is not wholly dispassionate and without reference to my own interests. Nor, I might add, is it congenital. As already noted, I was born an invertebrate neurophysiologist and, 10 years ago, had no more interest in lesion experiments than any other self-respecting invertebrate neurophysiologist. Indeed, it was in the course of studies on an honorary invertebrate nervous system problem, the nature of the circuitry underlying prey orienting movement in the frog, that I and my colleagues first stumbled on to the lesion methodology (Grobstein et al. 1978; Comer and Grobstein 1978). That we have continued to use it actively (see chapter 19) provides both an explanation for my interest in defending the methodology and part of the basis for such a defense: it works.

One other point is worth making in this somewhat autobiographical context: that we have found the lesion methodology useful not in a particulary complex situation but rather in a "simple system," one fully amenable to attack at the cellular and subcellular levels (cf. Ewert 1987; Ingle 1983). Kolb and Whishaw point out that the decline in reputation of the "macrotechniques of neuroscience" was associated with the rise of the "microtechnology of the 1970's and 1980's." What our experiences suggest is that unrealizable expectations raised by the latter have more to do with the current disrepute of lesion studies than does their own intrinsic limitations. The issue is whether more modern techniques actually superseded the lesion methodology, as opposed to triggering a reformulation of questions in such a way as to mask the usefulness of lesion experiments. This possibility seems additionally worth exploring, given the current sense that there are severe problems with the microtechnological perspective.

In considering the matter, it is worth reflecting on the origins of some important and widely used concepts in modern neurobiology. Sensory mapping, the notion that there is an orderly relation between place in the nervous system and place in a sensorium, was unequivocably established by lesion methods prior to the development of microelectrodes or of adequately high-resolution neuroanatomical techniques (Holmes 1945; Teuber et al. 1960). Central pattern generation, the capacity of the nervous system to endogenously generate complex spatio-temporal discharge patterns that underlie movement, was similarly established by lesion experiments (Wilson 1966; Grillner 1981). The latter is particularly interesting, since it is difficult to imagine ways in which the concept could have become established without lesion experiments. It is not at all clear that it would have emerged from single unit recording and tracing of connections. Even today, knowing that a pattern generating capacity exists in small groups of neurons, it has proven difficult to say what role any given neuron plays (Selverston 1980). Central pattern generation is by no means the only concept with an obligatory origin in lesion experiments. Functional brain lateralization, for example, was established by lesion experiments (Sperry 1974; see Oppenheimer 1977 for references to earlier literature); it seems highly unlikely it would have emerged from a systematic study of neurons and their connections. The same holds for, among other concepts, the "two visual systems" notion, and the increasing evidence for dissociation of processes underlying "seeing," "knowing," and "knowing that one knows," an early form of which attracted Marr's attention (see Weiskrantz 1986 for a recent discussion).

My point is not only that lesion experiments have provided many of the information processing concepts that have motivated subsequent microtechnological analysis, but, more importantly, that the lesion methodology is in fact appropriate not only to suggest hypotheses about meaingful units of brain organization but to prove such hypotheses. The demonstration of an orderly motoneuron discharge pattern from a nervous system in which all afferent pathways have been cut is a logically rigorous demonstration of the existence of central pattern generating circuitry (Grillner 1981). One can in fact go further and localize such circuitry to a significant degree: that the isolated mammalian spinal cord may display a locomotor pattern establishes the existence of pattern generating circuitry within this restricted piece of the nervous system. Indeed, in favorable cases, lesion experiments have established that a locomotor rhythm involves several distinct pattern generators, coordinated by efference copy mechanisms (Stein 1976). One could cite a host of similar examples (a more extended discussion of one appears in chapter 19 below), but the present ones seem to me adequate to establish that the lesion methodology is a good deal more potent than it is frequently given credit for being. Properly used, lesion experiments are adequate to identify information processing blocks relevant to particular behaviors, to localize such blocks to particular areas of neural tissue, and to characterize some of the relations among the blocks.

What lesion experiments have not, in general, been adequate to do is to establish the cellular mechanisms of the information processing blocks or their relations, and this has almost certainly been part of the reason for their existing poor reputation. With the ability to monitor cellular and subcellular processes has come a mindset that one really does not have an explanation of nervous system processes unless one has it at the cellular level. Central pattern generating circuitry may be a single neuron, a circuit of relatively simple neurons, or a complex interaction of circuitry and complex cellular properties. Similarly, a sensory map may or may not correspond to a topographic pattern of anatomical connections. What has been sacrificed in this insistance on reductionism is an awareness of the capability to ask meaningful questions at the level of the information processing blocks themselves. Questions such as how many pattern generators there are for a locomotory movement, and what is the nature of the coupling between sensory maps and pattern generators (Grobstein 1988b, 1989, chapter 19 below), can be meaningfully posed and answered, irrespective of whether one can account for properties of the intermediate level elements at the cellular level. If additionally, as seems increasingly to be the case, easily observable isomorphisms between neuronal function and behavior are absent at the cellular level and become apparent only at higher levels of organization (Grobstein 1988a,c) the noncellular focus of lesion experiments may well emerge as an asset rather than a liability.

Actually, a focus above the cellular level, while perhaps valuable, is not an intrinsic characteristic of lesion methodology. Lesion experiments need not involve gross and hence poorly specified damage. Simple systems neurophysiologists have begun making active use of single cell lesion experiments in an effort to overcome the limitations of microtechnology (cf. Comer 1985). An interesting and instructive outcome of one such analysis was to establish that a well-characterized circuit that looked like it ought to be responsible for triggering escape behavior in fish and shown to be adequate to trigger escape behavior was not in fact necessary for such behavior (Eaton and DiDomenico 1985). The issue under consideration was whether the Mauthner cell, a large identified neuron, is appropriately characterized as a command cell, in the sense of being necessary and sufficient to support escape movements. Hyperpolarization of the neuron blocked escape movements in response to eighth nerve stimulation, suggesting "that loss of the cell will prevent the escape behavior when the animal is given a threatening environmental stimulus. Paradoxically, this is not the case. . . . Despite loss of the cell, the escape response still happened." It is at such a point that many neurobiologists throw up their hands and conclude that, as they suspected, lesion experiments are just fundamentally uninterpretable. It is this that I suspect is as much reponsible for the negative feelings about lesion experiments as their relative lack of a cellular focus.

In actuality, the escape behavior findings are not only perfectly interpretable but quite meaningful: they establish that either of two different circuits may mediate what investigators had presumed to be a single rather simple piece of behavior (Eaton and DiDomenico 1985). They not only prove the existence of a significant functional redundancy but, further, show that the role a particular cell plays in what had been thought to be a single behavioral act varies depending on the circumstance. The apparent uninterpretability of the lesion findings actually reflects a failure to appreciate the intrinsic logic of lesion experiments: it is possible to use them to prove that a circuit is adequate to support a behavior but not that it is necessary. Once this fundamental reality and some further implications are clearly understood, the lesion observations become not only interpretable but quite significant generally. Both redundancy and a substantial context dependence of the activity of individual neurons are, as will become clear in the following, increasingly apparent as common aspects of neuronal organization, which are important to the development of general principles and which also play a notable role in making the microtechnology approach problematic at best. For reasons that I will briefly discuss in the next section, the logic appropriate to the interpretation of lesion experiments is counterintuitive to many neurobiologists. It is this problem, rather than intrinsic limitations, that seems to me to be perhaps the most significant contributor to the current poor reputation of the lesion methodology, and that needs to be overcome for the methodology to fulfill its potential for contributing to understanding the information-processing characteristics of the nervous system. The following discussion of what one can and cannot legitimately conclude from lesion experiments will I hope both provide further evidence of their usefulness and contribute to clarifying the logic of their interpretation.

Lesion Studies: Notes for a Logic

"This series of experiments has yielded a good bit of information about what and where the memory trace is not. . . . Although the negative data do not provide a clear picture of the nature of the engram, they do establish limits within which concepts of its nature must be confined. . . ." (Lashley 1950)

"The lesion results are consistent with the involvement of the crossing isthmo-tectal projection.... We cannot of course totally rule out the possibility that the lesions produced the observed results by interrupting some unknown intertectal pathway, either directly or indirectly, or by removal of some necessary facilitating influence...." (Grobstein et al. 1978)

"It is also clear from the results reported here that tectal ablation does not abolish the ability of the frog to generate spatially organized prey acquisition behavior. This implies that pre-motor circuitry capable of elaborating such behavior is present outside the tectum." (Comer and Grobstein 1978)

"The (lesion) result indicates that tectum is not an essential 'final common path' for all sensory channels to gain access to motor programs responsible for prey acquisition responses." (Comer and Grobstein 1981)

The lesion findings indicate "that a given (superficial) tectal region is connected to pattern generating circuitry in such a way that activity in the region can potentially activate a variety of different output patterns." (Grobstein et al. 1983)

"The hemisection syndrome thus implies a form of organization which establishes a relation between one side of the brain and one side of behavioral space." (Kostyk and Grobstein 1987a)

"Collectively, these (lesion) findings suggest that between sensory input and motor output there may be not only, as indicated in our model, topographic sensory maps and output pattern generating circuitry but an additional intermediate processing level which establishes a generalized spatial coordinate frame within which stimuli are located." (Kostyk and Grobstein 1987b)

From a naive perspective, the theory of lesion experiments is relatively simple: to determine what a part of the nervous system is doing, one removes the selected structure and defines its function in terms of the resulting behavioral deficits. There are a host of problems with this perspective, both practical and conceptual. Not the least of these is that such a theory presumes what need not in principle be the case: that the nervous system is organized in such a way as to display a high degree of localization of functions, with the localizable functions corresponding to known aspects of behavior. That many information-processing devices display neither of these characteristics is obvious to anyone having the most casual familiarity with computers. What is perhaps more important in the present context is that it is demonstrably not the case that one can safely presume these characteristics in the nervous system.

Given the history of a close linkage between lesion methodology and the localization problem (cf. Luria 1980, chapter 1) an adequate theory for lesion experiments necessarily starts with what lesion experiments *cannot* do. With this in hand, it becomes clearer what they can do. In the following, I will try and summarize our 10 years or so of experience with the lesion methodology in terms of a series of principles for the interpretation of such experiments. In doing so, I am acutely aware of standing on the shoulders of others, and it is by no means my intent to suggest that the principles outlined are either novel or exhaustive. As indicated in the previous section, and exemplified by the opening quotation of this one, the intrinsic logic of lesion experiments has been clearly appreciated by a number of investigators who have used it with great success for current understanding of the nervous system. Nor am I the only one to attempt a theory of interpretation for lesion experiments (cf. Dean 1982, for a recent treatment with conclusions not dissimilar from my own, and earlier references). If there is anything unique about the present statement of principles, it is perhaps that they have derived from experiences with a simple rather than with a complex system. This should make the relevance of a theory of lesion experiments more apparent to a larger number of neurobiologists and theoreticians. At the same time, there are a number of problems in the use of lesion experiments for exploration of more complex problems (cf. Humphreys and Riddoch 1987) that I will not discuss, at least not directly. I believe the principles to be outlined of relevance in all contexts, as I will try and make clear, but the list is almost certainly incomplete. My hope is that the explicit statement of some basic principles will not only make it clear that a useful and rigorous logic for the interpretation of lesion experiments exists, but contribute to the further explicit development of such a logic.

One Cannot Conclude from the Absence of a Piece of Behavior following a Lesion That the Structure Removed Is Essential for That Behavior

As implied by my earlier remarks, this principle is reachable in a number of different ways, and ought to be regarded as well established. It is nonetheless not infrequently rediscovered with some surprise (Eaton and DiDomenico

1985), and even among investigators experienced with lesion methdologies, it is a principle most frequently honored in the breach. A phenomenon originally described by James Sprague (1966; see also Sherman 1974, 1977) provides perhaps the most dramatic justification for the principle in terms specific to the nervous system; the papers should be required reading for anyone interested in the use of lesion experiments. The basic phenomena are simple and robust. Removal of visual cortex on one side of the cat brain results in a failure to orient toward visual stimuli in the contralateral visual hemifield, a finding straightforwardly interpreted as a dependence of visual orienting on cortical structures, and one that contributed to the generally held notice of an increasing encephalization of function in verebrate phylogeny. Sprague's noteworthy discovery was that visual orienting is reinstated by what turns out to be any of several subsequent midbrain lesions. The nature of the interactions responsible for the Sprague effect remains unclear. What is quite clear however is that the original deficit cannot be attributed to the removal of circuitry *essential* for visually elicited orienting; circuitry *adequate* to support such behavior survives the cortical lesion.

The distinction between "essential" and "adequate" circuitry is an important one for the theory of lesion experiments. It turns out to help clarify not only what they cannot establish, but what they can, as will be discussed further below. Before turning to that, however, some additional remarks about the nature of the constraints on the interpretation of the absence of particular behaviors following lesions ought to be made. In the example of escape behavior described earlier, inactivation of a particular neuron blocks a behavior when triggered one way but apparently not when triggered in a different way. That the nervous system has two different circuits to do roughly the same thing precludes characterizing either as essential for the behavioral task as defined. In the case of the Sprague effect observations, something different and more subtle is going on. Whatever the details of the circuitry underlying the Sprague effect, it seems fairly clear that damage to one part of the nervous system (the cortex) is altering the characteristics of a distant part (probably the midbrain) so that circuitry adequate to form a particular function is not expressed in behavior. A subsequent second lesion reinstates such expression.

The disappearance of behavior despite the persistance of circuitry adequate to support that behavior and serious misinterpretations of neuronal organization consequent on a failure to entertain this possibility are by no means phenomena restricted to studies of the sensory side of the nervous system. As late as 1974, a major medical textbook (Mountcastle 1974) noted: "... the bulbospinal cat ... cannot ... run, walk, or jump. That these deficits are due to the fact that the neural mechanisms essential for these more complex acts are situated rostrally to the bulbar region and not to some kind of 'shock' has been shown by experiments in which bulbospinal animals have been kept alive for considerable lengths of time." It is true that a bulbospinal animal does not walk. It is not true, however, that a bulbospinal animal lacks the neural mechanisms essential for this "more complex" act. Not only a bulbospinal animal but also a spinal one will exhibit locomotory limb movements if subjected to any of several kinds of generalized excitatory stimuli. In a similar vein, Lashley (1951), in what is perhaps still the most insightful paper on the interdependence of neuronal circuits, noted "A monkey, for example, after ablation of the precentral gyrus may seem unable to use the arm at all, but if emotional excitement is raised above a certain level, the arm is freely used. As soon as the excitement dies down, the arm is again hemiplegic. I have seen something of the same sort in a human hemiplegic."

The Sprague effect and paralytic phenomena in general provide modern examples of a body of phenomena well recognized by clinicians and physiologists in the late nineteenth and early twentieth centuries who observed and attempted to account for recovery of function after brain lesions (Rosner 1974). The phenomena, which went under a variety of names including diaschesis, have been of persistent interest among Soviet investigators (cf. Luria 1980) and may well be related to modern Western findings of alteration in cortical maps following peripheral lesions (cf. Merzenich et al. 1984). Such "remote effects" are almost certainly a much more common aspect of brain function than has been evident from microtechnological neuroscience, with its emphasis on detailed study of isolated parts of the brain, and worth increased analysis in the information processing context.

To try and make my point as sharply and generally as possible, I have chosen my examples to illustrate the need for the first principle of lesion experiments in terms of phenomena that are at least in principle relatively easy to account for in terms of known principles of nervous system organization, and where the behaviors to be accounted for appear, at first look at least, to be similarly straightforward. When a behavior is less well defined, the problems obviously magnify. One example will serve to make the general point. "Vision" in humans was for an extended period regarded as almost entirely a cortical function, based on the finding that humans following cortical lesions fail to report the existence of stimuli in defined regions of their visual field. It is now clear that vision is not a unitary function: a substantial, reasonably sophisticated visual capacity can be documented following cortical lesions if the task is defined properly (Weiskrantz 1986). What is lost is the capacity to report detection of visual stimuli. The problem is in some respects not dissimilar from that of escape behavior. What may appear as a unitary piece of behavior to an investigator is actually a number of different processes from the perspective of the nervous system. The upshot is that a piece of behavior may appear to be missing following a lesion when tested in one way, and not be missing when tested in another.

The general message is that lesion experiments cannot be used to prove that a particular piece of the brain is necessary for a particular piece of behavior or a particular behavioral task. It should be stressed that while I have given a list of phenomena that precludes the use of lesion experiments for this purpose, the problem is neither these particular phenomena nor any currently definable longer list of phenomena that could be systematically excluded as alternate explanations for the disappearance of a piece of behavior. The problem is more general: the nervous system displays a level of interconnectedness and interdependence that makes it impossible to presume that even a well-defined missing function after a lesion is attributable to removal of circuitry necessary for that function. We do not as yet have anywhere near the kind of understanding of either neuronal organization or of information processing that would exclude all possible expanations leaving only the desired interpretation: that absence of behavior reflects removal of essential circuitry. (See Grobstein 1988c for a more general discussion.)

This may seem a harsh conclusion but it seems to me logically inescapable, and one that must be clearly acknowledged if a theory of lesion experiments is to be placed on a sufficiently sound footing to overcome the current disrepute of the technology. Needless to say, I do not regard the inability to prove necessity of a neural structure for a behavior from the absence of behaviors following a lesion as a blanket indictment of the lesion methodology. Indeed, acknowledgment of the limitati
itself yields important general insights into nervo
system organization, as discussed, and also provides
basis for defining the sorts of more positive conclusio
that can be reached from lesion experiments, as describ
below. It is also worth noting that similar reservatio
about establishing causality, though not often acknow
edged, hold for microtechnological approaches to bra
function as well. They also hold for frequently used lesio
like experiments in other areas of science; indeed, as di
cussed below, it is arguably the case that most scienti
observations actually establish causal adequacy rather th
necessity; this capability of the lesion methodology
considered next. Finally, it is worth noting that determi
ing whether an aspect of behavior disappears following
brain lesion can certainly be meaningful in particular we
defined circumstances. Given an hypothesis about t
role of a particular brain structure developed in terms
some other line of evidence, the predicted loss of
aspect of behavior may provide valuable supporti
evidence for the hypothesis; the failure to observe t
predicted loss may invalidate the hypothesis, if it
properly phrased. The latter is discussed further below,
the use of deficits to generate hypotheses.

One Can Legitimately Conclude from the Persistence of Behavior following a Lesion That the Surviving Circuitry Is Adequate to Support That Behavior

This principle is in a sense, but with an important qual
fication, the converse of the first principle, and may see
both obvious and trivial. In fact, once appreciated, it is
enormously valuable asset in characterizing the function
organization of the nervous system. I have already,
several points in the preceding, provided examples of i
use. Awareness of the existence and location of circuit
adequate to generate complex spatiotemporal motoneuro
discharge patterns that underlie movement emerged fu
damentally from application of the second principle. Th
removal of structures rostral of the spinal cord as well
of all afferent input spares the capacity of the spinal co
to generate appropriately patterned motoneuron di
charge patterns is a clear and rigorous documentation
the existence of locomotory pattern generating capabilit
in the spinal cord. Similarly, the persistence in humans
an ability to point toward visual targets following cortic
damage (Weiskrantz 1986) is a straightforward proof th

noncortical circuitry is adequate to support that function, subject only to technical reservations about the adequacy of the lesions and testing paradigm.

The second principle may seem to lack general applicability, the examples mentioned being in some sense special cases. I think this is actually a function of the fact that (so far as I know) the principle has never been clearly stated as a general one, and that it seems backward from the way one normally thinks of lesion experiments. Indeed, an appreciation of the principle in my laboratory grew from what was initially perceived as a failed experiment. As is discussed in chapter 19, we were intrigued by the existing evidence for polymodal convergence in the vertebrate optic tectum, and believed we could establish the functional significance of such an arrangement by making small tectal lesions in the frog and documenting corresponding regions of visual and nonvisual sensory space (tactile in the case investigated) where stimuli failed to elicit orienting responses. No one then (or since) has documented such corresponding sensory scotomas, and we were unable to find them. Finally, in frustration, we removed the entire optic tectum, with the result that visual orienting was, as expected, abolished, but tactually elicited orienting persisted (Comer and Grobstein 1978, 1981). It was only after thinking the matter over a bit that we realized that instead of having supportive evidence (the first principle) for what we expected, we had conclusive evidence for what we did not expect.

Once made explicit, we realized that the second principle can be used systematically, not infrequently in ways that overcome limitations imposed by the first principle. Anatomical evidence for a crossed tectospinal projection suggested that the neuronal organizatio.ı underlying visually elicited orienting in the frog might ınvolve two sequential decussations, the first in the retinal projection and the second in the tectofugal pathway, However, uncrossed descending tectofugal projections, though less dramatic, also exist, so that the anatomy would be equally consistent with the involvement in orienting of an uncrossed descending tectofugal pathway. Following observations of Sperry in the newt, we performed lesion experiments in the frog that provided definitive evidence that there exists a crossed descending tectofugal pathway adequate to support orienting turns (Kostyk and Grobstein 1982, 1987a). The experiments reflected a direct application of the second principle, and consisted of documenting the persistence of visually triggered orienting turns toward stimuli represented in the contralateral tectal lobe following complete interruption of all uncrossed descending projections from that tectal lobe (see figures 1 and 2 in chapter 19). It is worth making explicit that this apparently backward form of lesion experiment yields a definitive finding while the normal form, observing an absence of turns following interruption of the crossed projections from the tectal lobe (Ingle 1983), did not.

The experiments described, in addition to verifying the existence of an expected pathway, a crossed one from each tectal lobe adequate to trigger turns in a direction contralateral to that tectal lobe, also suggested the existence of an unexpected one, an uncrossed pathway from each tectal lobe adequate to trigger ipsilateral turns (see figure 1 in chapter 19). Following complete interruption of descending projections on one side of the brain, frogs failed to turn toward stimuli at any location in the entire ipsilateral visual hemifield. A failure to turn toward stimuli in the monocular visual field is expected, given that the lesion interrupts a crossed projection from the opposite tectal lobe. Binocular visual field, however, is functionally represented in the frog in both tectal lobes. The failure to turn toward stimuli in the ipsilateral half of binocular field would be simply accounted for if the lesion interrupted, in addition to the crossed projection from one tectal lobe, an uncrossed projection from the other. The existence of this pathway could not, however, according to the first principle of lesion experiments, be inferred from the deficit. Appropriately designed experiments based on the second principle of lesion experiments got us around this constraint. By interrupting all crossed projections and showing a persistence of orienting turns in the appropriate direction, it was possible to establish, again rigorously, the existence of the unexpected uncrossed path adequate to trigger ipsilateral turns (Kostyk and Grobstein 1987b).

I have described our own use of the second principle in some detail not because the findings are of special significance (though they have proven quite useful in further work on the nature of spatial representation in the frog brain; see chapter 19 below) but because it seems to me the principle itself is underappreciated and hence underused. A number of existing weak inferences about the involvement of particular structures in particular behaviors based on deficits might be substantially strengthened by demonstrations of the persistence of behavior following damage to other structures. This strategy is actually a generalization of a double dissociation protocol, familiar to some physiological psychologists but probably not to most neurobiologists. The principle can also, as I hope I have illustrated, be used in its own right to make quite meaningful statements about information-processing capabilities of structures in the nervous system and about information flow between them.

Lesion experiments based on the second principle can be extremely powerful in characterizing information-processing characteristics of the brain, but I would be remiss if I did not also point out some of their limitations. The technical problem of being certain of what has been removed is obvious. At the same time, one has to worry less about what may or may not have been indirectly damaged by the lesion than in the case of trying to attribute particular deficits to particular locations of disturbance, since one is interested in survival of behavior after removal of a particular structure rather than its disappearance. What may be a less obvious technical requirement associated with the second principle is that behavior must be tested sufficiently soon after the lesion to provide reasonable assurance that surviving behavior is attributable to original circuitry that survived the lesion, rather than to a possible reorganization in surviving circuitry. There is of necessity a certain amount of ambiguity on this point, but the issue is fairly clear in principle, and frequently in practice. An additional more or less technical problem with the second principle is that it puts one in the position of designing experiments in which only one of several possible outcomes (survival of a piece of behavior) is directly interpretable. In practice, however, alternate outcomes (disappearance of behavior, alteration of behavior), though not directly interpretable, are frequently meaningful in other ways, as discussed further below.

A more serious conceptual limitation is implicit in the deliberate wording of the second principle: what can be demonstrated is *adequacy*, not *necessity*. The demonstration that a piece of neural tissue can support a particular piece of behavior is not equivalent to the demonstration that it is necessary for that piece of behavior. I have noted already that the nervous system frequently exhibits multiple ways to accomplish what the investigator may regard as a single piece of behavior. That the spinal cord contains pattern generators for locomotion does not preclude the possibility that such circuitry also exists in more rostral brain structures. Fortunately, this sort of problem, once recognized, can be at least further explored using the second principle: what is required are lesions appropriate to show that either of two different sets of neural structures suffices to support the behavior in which an investigator is interested. A more general aspect of the second principle's limitation to a demonstration of adequacy is that it precludes a rigorous conclusion about the role of the investigated structure in normal behavior. Strictly speaking, all that can be shown is the adequacy of surviving circuitry in the context of the damaged brain. Whether the circuitry functions in the same fashion, or indeed at all, in the intact brain is a distinct question, requiring other methods for its investigation, such as chronic recording of other forms of relatively noninvasive monitoring of brain activity. Again, the limitation needs to be acknowledged but more frequently than not can be reasonably effectively dealt with in terms of other kinds of observations (see chapter 19 below for some examples). It is certainly a less serious limitation than those associated with many microtechnological observations, in that at least adequacy can be assured. Finally, as I have already suggested and will consider further below, "necessity" may in fact be a poor criterion not only for brain research but for science in general (see also Grobstein 1988c).

A final point that should perhaps be made is that the use of the second principle of lesion experiments can in principle proceed as a data-collecting exercise but is best done from some theoretical basis. Clearly, establishing that a small lesion of the visual cortex does not affect beating of the heart is not a particularly valuable piece of information (though in fact cataloguing the areas of the brain which when damaged do not affect heart beat was of substantial importance in the history of investigations of brain function [Jeannerod 1985]). With some acknowledged expectation, however, documenting persistence of a function after brain damage can be quite significant.

One Can Legitimately Conclude from the Behaviorial Abnormalities following a Lesion That the Brain Is Organized in Such a Way That a Disturbance of Its Oraganization Would Have That Particular Behavioral Consepuence

Just as the second principle may initially appear obvious and trivial, the third is likely to strike one as contorted and perhaps tautological. The key to understanding the principle, and the source of its power, is that there are in general a number of different ways the brain might be organized in relation to particular behaviors, and they frequently differ in the possible abnormalities that could result from disruption of that organization. A careful characterization of the behavioral sequelae of disruption can frequently eliminate whole classes of hypothetical forms of brain organization, and imply others.

What may have been the most generally important use of the third principle in neurobiology was to establish the principle of functional heterogeneity in the cortex (cf. Luria 1980). That the behavioral consequences of lesions at different cortical locations are different eliminates the possibility that, for example, the cortex is a general purpose, distributed processor or information storage system. In so doing, lesion experiments provided the basic rationale for virtually all subsequent work on the cortex. The example is also instructive as an illustration of what the third principle does *not* allow. Just as observations of deficits following lesions is a less interpretable outcome in connection with the second principle, so the failure to observe abnormalities is a less interpretable outcome with the third. Lashley's classic observations, for example, that cortical regions are equipotential with regard to learning (Lashley 1950) clearly reflected a phrasing of the behavioral question in such a way that existing functional heterogeneity in the cortex was not made evident. At the same time, the observations certainly did eliminate a number of conceivable learning models based on specific intracortical pathways.

A more general constraint in connection with the third principle is that it allows a rigorous statement of what an organization is *not* but not of what it is. That cortical regions differ from one another can be stated with certainty. That one is reponsible, for example, for vision and another for poetry cannot appropriately be concluded, since that would represent only one of an unknown and perhaps unknowable number of forms of organization that would yield a particular observed set of behavioral abnormalities following particular brain lesions. In short, while functional heterogeneity can be established by brain lesions using the third principle, no particular form of functional localization can be. That is not to say that testable hypotheses cannot be derived, as discussed at greater length in the next section, but only that proof is not obtainable in this way.

The phenomena of functional brain asymmetry, mentioned earlier as one of the organizational principles unlikely to have been deduced except via lesions, can be thought of as a special case of cortical heterogeneity, and as such has the same logical status. That the organization of the brain is such that damage to the same structure on one side has a different effect than to that on the other is certain; that this is because one side is specialized for one set of tasks and the other for another is an inference that requires proof in other ways. The blind sight phenomena, also discussed earlier, is interpretable not only in terms of the second principle of lesions (subcortical circuitry is adequate to support reasonably complex visuomotor function), but also the third. It clearly establishes that "awareness" is not a prerequisite for active sensorimotor processing, a point that, once stated, seems obvious in terms of day to day experience, but that is much less clearly recognized in the implicit models that underly most investigations of brain organization. One might add that the inverse is almost certainly also true: active sensorimotor processing is not a prerequisite for "awareness," as evidenced by the vividness of dreams. Clearly, hierarchical models of brain organization, in which awareness occupies an obligatory position at the top and has a role restricted to the control of active input/output relations, can be excluded. Similarly, as noted by Marr, lesion observations imply that information about object shape can reach awareness without obilgatory passage through a stage of recognition in terms of use and purpose, and hence eliminate information-processing models in which the latter necessarily precedes the former.

Though clearly relevant for this purpose, I do not mean to give the impression that problems of "higher brain function" are the only ones for which the third principle is relevant. Indeed, the principle in its present formulation, like the other principles, became evident to us as a further result of our experiences in trying to understand the nature of the linkage between the retinotectal map and pattern generating circuitry in the case of orienting behavior in the frog. As described at greater length in chapter 19 (see also Grobstein 1988b), we were struck in the course of a series of lesion studies by the repeated observation that damage to various parts of the brain did not eliminate the triggering of movements by selected regions of the superficial retinotectal projection but rather changed in systematic ways the particular movements tiggered. That the neuronal organization underlying orienting was such as to make this outcome possible was by no means a foregone conclusion. Indeed the prevalent theories at the time to explain orienting behavior involved a presumption that the nervous system was organized so as to create a one-to-one correspondence between location in the superficial retinotectal projection and movement. That each tectal region can activate any of a number of responses clearly eliminated a large class of models for describing the neuronal organization between tectum and motor output, and contributed to the development of a new one (chapter 19). A noteworthy characteristic of the latter is the recognition that remote effects, of the kind mentioned earlier, are critical in accounting for the linkage between tectum and motor output. Not only network structure but on-going patterns of activity determine the nature of the linkage at any given time. This characteristic has been termed "activity-gated divergence" (Grobstein 1988b, 1989) and is discussed at greater length in Chapter 19.

Similar alterations in understanding of the information processing underlying orienting resulted from a second phenomenon repeatedly encountered in the course of our lesion studies on the frog: that deficits resulting from lesions at locations caudal to the tectum have a geometric character to them, typically relating specifically to one of the three axes of stimulus location and having a border on the midsaggital plane. Again, as detailed in Chapter 19, it was possible to draw a rigorous conclusion about brain organization, that there is a relation between laterality of brain and laterality of behavioral space, a conclusion that was meaningful in its own right and surprising in terms of any available evidence. It clearly excluded a class of simple models for describing the linkage between tectum and movement. It also suggested some new models, involving a previously unsuspected intervening form of spatial representation (see chapter 19).

Here, as in the preceding section, my concern in describing particular experimental observations is to try and illustrate the usefulness and generality of a mode of interpretation of lesion experiments that may appear counterintuitive but is in fact logically rigorous and productive. The third principle, like the second, is most obviously applicable in the context of an existing hypothesis, and provides a basis for disproving it. At the same time, like the second principle, the third can be used agnostically. Behavioral abnormalities following brain lesions frequently can be used to eliminate whole classes of relevant if not yet explicitly entertained hypotheses, once one gets used to thinking of lesion methodology in these terms. More importantly, perhaps, the third principle, unlike the second, does not have the problem of uninterpretable outcomes. An observed behavioral abnormality following a brain lesion is a property of the system under investigation, and necessarily provides evidence about the organization of that system. It not only cannot be of such a sort that that abnormality would not occur but must be of such a sort that it can. The process of finding the question to which one has the answer is, as recognized by Marr, frequently an enormously productive way to advance the characterization of neuronal organization. Here too, however, as with the second principle, some findings obtained in lieu of a theoretical foundation are likely to be more significant than others.

One Can Legitimately Use the Results of Lesion Experiments to Motivate and Test Hypotheses about Information Processing Blocks and Their Interrelations

In outlining the first three principles of lesion experiments, my concern has been not with the more straightforward problem of how lesion experiments can be usefully employed but rather with the more formal problem of how they can be used (and not used) to make logically coherent statements about neuronal information processing and organization in relation to behavior. In this, I have been applying an unusual high standard to the methodology, one rarely applied in the case of most microtechnologies with regard to their relevance for similar objectives. Measuring lesion experiments against this high standard seems appropriate given the current poor reputation of the methodology, a reputation which is at least in part the result of assertions by investigators that the methods are capable of establishing conclusions (which they are not). At the same time, scientific methodologies are normally evaluated not in terms of their a priori logical status vis-à-vis some set of problems but rather in terms of their staying power in terms of suggesting hypotheses and resolving questions of interest to investigators. In these terms, the lesion methodology clearly needs no apology. With the possible exception of pharmacological manipulations, no procedure for exploring nervous system organization has a comparably long and successful track record. The success of that record is probably attributable at least as much to what I would call "bootstrapping" as it is to strict attention to logical rigor. The present discussion of lesion methodology would be incomplete if this aspect of its use were not at least mentioned.

Science is frequently characterized as a process of enumerating possible explanations of a phenomenon, and then designing an experiment that establishes that one of the possible explanations is in fact the case. In actuality, things are rarely (perhaps never) so neat. The enumeration of possible explanations sounds straightforward, but it is actually an intrinsically mysterious step, a characterization of which would itself be a major neurobiological accomplishment: where hypothetical explanations come from is perhaps the greatest mystery of brain function (James 1890). Because of this uncertainty, it is always possible that there are explanations other than those enumerated, and further possible that an experimental outcome presumed to reflect a particular explanation actually reflects a totally different one. The uncertainties increase with the complexity of the phenomenon being explored, and clearly are quite significant in dealing with the phenomena of neuronal information processing. This sort of problem is normally dealt with by appeal to some form of Occam's razor argument: one entertains only the simplest explanations or the most elegant ones. Beyond the obvious fact that these terms are themselves mysterious and poorly defined, it seems clear that their use has frequently been misleading in explorations of brain organization. To cite but one example, the phenomena of substantial redundancy in the neuronal circuitry underlying particular behaviors clearly proved incompatible with what seemed at the time reasonable dictates of Occam's razor.

The upshot is that studies of brain organization dramatize the need for something in addition to a first-principles, deductive research strategy. What is required is a more inductive approach, in which the motivation of hypotheses by observations is as important as the motivation of observations by hypotheses (see Grobstein 1988a). With such an approach, the logical desideratum is one of seeking sense rather than proof. Hypotheses emerge as one detects patterns in the observations made. The hypotheses in turn motivate new observations to determine whether the perceived patterns hold in new contexts. As new observations are made, the patterns inevitably alter, leading to hypotheses summarizing increasingly wider ranges of observations, and so on. One important characteristic of such an inductive approach is that it has a built in error-correcting mechanism. By testing patterns in novel contexts, rather than trying to verify them in simpler ones, one avoids problems of inadequate or fundamentally wrong first principles. An equally important characteristic of the inductive approach is that it frees one from the reductionist fallacy (Grobstein 1987a) of believing that one has to know the properties of elements of a complex system before one can proceed to usefully characterize the system itself. Patterns in observations can be detected at any level of organization with which one chooses to work.

From this perspective, the key question with regard to the lesion methodology is not so much what it can or cannot prove, as how good it is, relative to other methodologies, at generating an expanding cycle of hypothesis and observation relevant to understanding neuronal information processing. As implied already, the answer to this question seems fairly clear and positive. The localization problem itself provides a good example in this regard (Luria 1980). Initial observations implying discrete localization of function led to further observations in new situations that proved less consistent with discrete localization. Those in turn led to more sophisticated concepts of what is meant by function, and to further and more subtle observations that have clearly contributed to the progressive unraveling of what is involved in various forms of abstract information processing. Our own experience, in a more specialized context, has been quite similar. What began as a relatively straightfoward effort to gather evidence consistent with a map-like coupling between a sensory map and pattern-generating circuitry has now gone through several stages of hypothesis and refutation (Grobstein, this volume). The most recent observations suggest the existence of an intermediate representation of spatial location in an unexpected and rather abstract coordinate frame (Grobstein 1988b), as discussed at length in Chapter 19. While the lesion observations themselves are not adequate to prove that a representation of exactly the currently hypothesized form exists, they are more than adequate questions that point to new situations in which to collect information. The ability to repeatedly predict otherwise arbitrary findings, as for example the existence of an uncrossed pathway adequate to trigger ipsilateral turns, and a consistency of implication across a large number of situations, as for example the repeated finding of generalized alterations in sensorimotor linkages following lesions, may well be a more useful form of proof in scientific inquiry than any deductive process.

Clearly, the general inductive or bootstrapping strategy is not unique to the lesion methodology. It can be used for purely behavioral studies, for studies of neuronal structure at any level of organization one wishes, and, indeed, for any microtechnological methodology as well. What is distinctive about the lesion methodology is its presumption of isomorphisms between brain organization and behavioral or information-processing organization without presumption as to the level at which such isomorphisms must exist. What this yields is an hypothesis-generating methodology uniquely fitted to generating descriptions of elements relevant to describing both brain organization and the information processing that underlies behavior. Whether lateralization in humans or a space map in frogs has simple correlates in terms of nervous system organization remains to be determined. That they must have some correlate could probably not have been established except by way of lesion experiments.

Lesion Studies: Present and Future

In the introduction to this chapter, I asserted that studies not only provide a logically coherent way to approach the problem of analyzing the information-processing characteristics of the brain, but further that they may well provide not only a more successful analysis than that resulting from many microtechnological approaches but one needed even given a combination of microtechnology and computational theory. The argument is partly the increasing awareness of the limitations of microtechnological observation, coupled with the successes of lesion studies, as considered in preceding sections. There is however a more general argument, alluded to in the introduction. For decades, it has been an article of faith that the neuron is the basic element of neuronal information processing, as it is, by and large, of the nervous system as a structural entity. Implicit in the presumption is that there are definable elements of the information processing underlying behavior that can be identified with the connections and behavior of individual neurons, and hence that a bottom-up approach will lead progressively to an understanding of neuronal information processing at higher and higher levels of complexity. Experiences with microtechnological approaches have made this an increasingly untenable starting point, and suggested the necessity for a top-down, computational approach instead. With an eye on the future and an objective of developing better strategies, its worth trying to specify as explicitly as possible what the problem has been, and why a third, intermediate level approach is necessary. (See Grobstein 1988c for a more general discussion.)

In one of the best known papers reporting the results of lesion studies, Lashley (1950) wrote "I sometimes feel, in reviewing the evidence on the localization of the memory trace, that the necessary conclusion is that learning is just not possible." Though frequently ignored in the modern literature, Lashley's analysis was a paradigmatic application of logically rigorous lesion experiments: "Although the negative data do not provide a clear picture of the nature of the engram, they do establish limits within which concepts of its nature must be defined." The paper, together with a companion (Lashley 1951), also anticipates a number of significant generalizations from lesion experiments, which account for what has emerged as problems with the microtechnological perspective.

The modern recognition that behavior is often a good deal more complex than investigators of the nervous system give it credit for being was clearly foreshadowed by Lashley: "Much of learning theory has been based upon supposedly isolated and simple instances of association, on the assumption that these represent a primitive prototype of memory. However, an analysis of even the conditioned reflex indicates that it is not the simple, direct association of stimulus and response that it has been thought to be (Lashley 1950)." Corresponding to this were a series of further insights into neuronal organization and its relation to behavior that have a remarkably contemporary flavor. Progressive degradation of behavior with lesion size, as opposed to discrete loss with particular restricted lesions, indicates the existence of some kind of distributed processing, as opposed to discrete localization of unitary behavioral functions. Behavioral deficits following brain lesions are frequently reversible, by alteration of input, context, or additional brain lesions, implying a substantial degree of interaction between the distributed processing elements." ... [I]nput is never into a quiescent or static system, but always into a system which is already actively excited and organized.... Only when we can state the general characteristics of this background of excitation can we understand the effects of a given input." (Lashley 1951). Both of these points, which follow logically from lesion findings known to Lashley and developed since (as discussed in previous sections), imply a kind of information-processing oraganization in the nervous system organization more like that characteristic of modern inquiries into computer science than that implicit in the microtechnological perspective. Moreover, distributed, interactive processing, as considered in the examples here, has proven to be a characteristic not only of complex systems but of relatively simple ones as well. Here too, Lashley displayed a remarkable prescience: "I have come more and more to the conviction that the rudiments of every human behavioral mechanism will be found far down in the evolutionary scale.... If there there exist, in human cerebral action, processes which seem fundamentally different or inexplicable in terms of our present construct of the elementary physiology of integration, then it is probable that that construct is incomplete or mistaken, even for the levels of behavior to which it is applied." (Lashley 1951)

Both the successes of lesion experiments and the frustrations of microtechnology point to the same conclusion: the neuron is not the optimal unit for analysis if one is interested in the information-processing characteristics of the brain. In this regard, a more general discussion of the relation between properties at different levels of organization in neural development and function (Grobstein, 1988a,c) is also germane. In those papers I drew attention to a similarity between recent findings on the relation between cellular recognition properties and neural networks, and those on the relation between neural networks and functional input/output relations. The conclusion, in both cases, was that specificity in organization at the more complex level was a function not of a corresponding specificity of some property at the less complex level, but rather of the interaction of several less specific processes at that level. A given set of recognition processes will support any of a range of neural networks, depending on, for example, how many neurons are present and what patterns of electrical activity have occurred in the network. Similarly, a given network will support any of a range of input/output relations, contingent on, among other things, the hormonal milieu, and the particular pattern of electrical activity in the network at the time a particular input occurs. What seems to be going on, I suggested, is that the properties of higher levels of nervous system organization are intrinsically not deducible from those of the elements because the former depends on information acquired since that which determines the properties of the elements. In short, because the nervous system is an information-*gathering* device the properties of its elements bear no fixed relation to the properties of the overall system (Grobstein 1988c).

The implication of the abstract argument is identical to that which has emerged in practice: one cannot in general deduce from an analysis of neuronal properties the information-processing characteristics of the neuronal ensembles that are ultimately expressed as behavior. Such properties may lead to hypotheses about significant neuronal ensembles but may equally well mislead one into believing that they are simpler than they actually are. It is because of the poor correspondence between neuronal properties and information-processing characteristics that Marr and others have argued for a "top-down" approach, one beginning with the computational problem to be solved, working through possible representations and alogrithms, and only at the end worrying about hardware implementation (the nervous system). The top-down approach is not unreasonable, given the perspectives of a computer scientist. On the other hand, a characterization of the computation to be accomplished is itself an hypothesis, the origins of which are obscure for reasons mentioned earlier, and, for the same reasons, one can never be sure one has entertained all possible representations and algorithms. Beginning with the computational task is fine if one is dealing with an engineering problem; it can be highly misleading if one is concerned with a preexisting complex information-processing device whose computational tasks, styles, and constraints are in fact part of what needs to be discovered. If one is interested in the nervous system and behavior, one might hope for some intermediate level approach: one free of the necessary presumptions of either the cellular neurophysiologist or the computer scientist and, ideally, one that can enrich the understandings of both.

My abstract argument provides a level of analysis characterization different from that of Marr, and perhaps one more instructive in the present context. What is of concern are levels of complexity of biological systems, rather than a sequence from more abstract to more concrete in the characterization of an information-processing device. The thrust of my earlier argument, and of this article more generally, is that somewhere between the level of neurons and that of the entire brain, there are levels of neuronal organization above the neuron at which the information-processing events whose interactions constitute behavior become more apparent. Central pattern generation is a discrete form of information processing that can be identified with reasonably localizable ensembles of neurons; the same is true of, for example, sensory maps and corollary discharge patterns. Such "information-processing blocks" provide good descriptors for both neural and behavioral organization. Appropriate studies of the brain can not only reveal such blocks but rigorously establish their existence and meaningfully explore the nature of the interactions among them. It is this kind of intermediate level brain analysis that seems to me necessary to bridge between cellular and global considerations, of either the neuronal or computational kind, and it is for this kind of intermediate level analysis that lesion experiments, used as discussed in preceding sections, are particularly well suited. The top-down computational approach risks misstating the point of the exercise. The intermediate approach is useful both in defining cellular problems to be solved, and in revealing computational problems and solutions about which the conscious theoretical brain may be totally ignorant.

Top-down and bottom-up analyses are familiar strategies for scientists dealing with complex systems; beginning in the middle is perhaps less so. A top-down analysis presumes that one can begin with what a system does and reduce this description successively to progressively simpler events, which ultimately are explainable in terms of the properties of a set of basic elements. A bottom-up analysis presumes one can begin with properties of the basic elements, explore interactions of units at increasing levels of complexity, and come ultimately to a full description of what the device accomplishes. If neither of these is ture, how is one to proceed? In this regard, it is worth noting that neither top-down nor bottom-up strategies are employed in what are almost certainly the most common and successful analyses of complex systems undertaken by human beings: the constructions of an understanding of reality by children. Children proceed neither from a fixed assumption of what reality is for not from a presumption of what elements make it up. Instead, they derive explorable hypotheses about both from the patterns of change created by their own actions, essentially from lesion experiments on the real world. It is this style that is of the essence of an intermediate level approach. That it should be effective for analyses of neuronal computation, as elsewhere, is no mystery (see also Grobstein 1988c).

Kennedy's Astrodome metaphor, with which I began this essay, was appropriate at a time when the merits of single unit versus global explorations of brain function was the point of debate, but needs some updating in the present context. Kennedy's interplanetary visitors displayed a remarkable restraint, restricting themselves to noninvasive methods of trying to understand the activity associated with a baseball game. While comendably humane (or extrahumane), the restraint condemned the investigators to working with what were at best uncertain correlations between observations and hypothetical forms of behavioral organization. A more successful analysis would seem to require that the investigators be willing to perturb the system under investigation, and so one ought to expand Kennedy's metaphor to allow the emergence of, in addition to mathematicians and microprobers, at least one additional group of investigators, whom one might term the "bootstrappers." This third group would of course initally be under heavy attack from the first two, if for no other reason than there is initially only one Astrodome to study and mucking around with it would seem not only highly unlikely to yield interpretable results but destructive of any possibility of more systematic study. Fortunately, as the mathematicians and microprobers collect useful but increasingly frustrating information, it becomes apparent that behavior similar to that observed in the Astrodome occurs in a number of other locations. Moral quandaries are partially if not fully resolved, and a sufficient number of distinguishable hypotheses have emerged so that the bootstrappers can justify perturbations that will both distinguish among existing hypotheses and generate new ones.

As work proceeds, it becomes clearer and clearer that the system under investigation is actually organized in such a way that perturbations are a more effective means of exploration than might have been thought. While the consequences of dropping large explosive devices into different locations turn out too similar and difficult to interpret, the sequelae of more focused perturbations prove more instructive. It turns out, for example, that, contrary to expectations of some of the microprobers, the overall pattern of activity is relatively resistant to removing any of the individual elements (each of which, it has been realized by this point, is a complex living system in its own right). The elements are however clearly not functionally equivalent, as hypothesized by some of the mathematicians: removal of a distinctively positioned group of six blue clad elements has more substantial effects than removal of a comparable number of the outlying varicolored elements. Other previously unsuspected functionally significant groupings of elements, at least two of them also color-coded, become apparent as the bootstrappers expand their inquiries, and begin to recognize which sorts of effects following removal are relatively specific and which are relatively nonspecific. New and ultimately productive hypotheses emerge, as it is realized, for example, that the previously described global property of "winning" is actually neither a property of all the elements taken together nor of any indivdual element but rather is associated with the two recently recognized color-coded groups.

My recasting of the Kennedy metaphor ignores an important aspect of the reality of terrestrial brain research, that lesion studies were an important aspect of such research from its inceptions, and perhaps somewhat inappropriately links computational neuroscience with global forms of investigation. Both computational and experimental neuroscience can be done and of course are being done at all levels of biological complexity; the salutary effect of computational neuroscience, however, has been largely to renew attention of all investigators to the problems at the higher levels of complexity. As for the seminal role of lesion studies in explorations of neuronal organization, I have taken pains to emphasize this point in the preceding, and it is with the present and future of the lesions methodology rather than the past that I am primarily concerned. For this point, the revised metaphor would seem to serve the purpose. The brain has clearly proven to be neither a sufficiently dedicated sort of computer so that one can deduce its properties from the properties and connections of individual elements, nor so general purpose a computer as to make its hardware irrelevant in determining its function. It has detectable and characterizable properties at an intermediate level of organization, what I have termed information-processing blocks. What I hope I have suceeded in establishing is that the characterization of these and their interactions is and will continue to be an important component of computational neuroscience, and that the lesion methodology is an important and valuable contributor to that analysis.

Acknowledgments

Supported by NIH Grant 1 R15 NS 24968 and a grant from Bryn Mawr College. This article reflects long and productive associations with a number of student and faculty colleagues, including Christopher Comer, Sandra Kostyk, Tom Masino, and Paul Patton. My thanks particularly to the latter two for reading versions of this manuscript and, more generally, for helping me to formalize principles of conduct that had developed out of day to day practice in our laboratory. My thanks also to a number of anonymous reviewers of other manuscripts and grant applications who provided me the incentive to formalize these principles.

References

Arbib MA (1975) Artificial intelligence and brain theory: unities and diversities. *Ann Biomed Eng* 3: 238–274.

Arbib MA (1985) Brain theory and cooperative computation. *Human Neurobiol* 4: 201–218.

Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1: 371–394.

Comer C (1985) Analyzing cockroach escape behavior with lesion of individual giant interneurons. *Brain Res* 335: 342–346.

Comer C, Grobstein P (1978) Prey acquisition in atectal frogs. *Brain Res* 153: 217–221.

Comer C, Grobstein P (1981) Involvement of midbrain structures in tactually and visually elicited prey acquistion behavior in the frog, *Rana pipiens.* J Comp Physiol 142: 151–160.

Davis, WJ (1976) Organizational concepts in the central motor networks of invertebrates. In Herman RM, Grillner S, Stein PSG, Stuart DG (eds), *Neural Control of Locomotion.* New York, Plenum, pp 265–292.

Dean P (1982) Analysis of visual behavior in monkeys with inferotemporal lesions. In Ingle D, Goodale M, Mansfield R (eds), *Analysis of Visual Behavior.* Cambridge, MIT Press, pp 587–628.

Eaton RC, DiDomenico R (1987) Command and the neural causation of behavior: A theoretical analysis of the necessity and sufficiency paradigm. *Brain Behav Evol* 27: 132–164.

Ewert J-P (1987) Neuroethology or releasing mechanisms: prey catching in toads. *Behav Brain Sci*, in press.

Grillner S (1981) Control of locomotion in bipeds, tetrapods, and fish. In Brooks V (ed), *Handbook of Physiology. Section 1. The Nervous System. Vol. 2.* Bethesda, American Physiological Society, pp 1179–1236.

Grobstein P (1987) The nervous system/behavior interface: Levels of organization and levels of approach. Commentary on target article. *Behav Sci* 10: 380–381.

Grobstein P (1988a) On beyond neuronal specificity: Problems in going from cells to networks and from networks to behavior. In Shinkman P. (ed), *Advances in Neural and Behavioral Development. Volume 3.* Ablex, Norwood, New Jersey.

Grobstein P (1988b) Between the retinotectal projection and directed movement: topography of a sensorimotor interface. *Brain Behav Evol* 31: 34–48.

Grobstein, P (1988c) From the head to the heart: Some thoughts on similarites between brain function and morphogenesis, and on their significance for research methodology and biological theory. *Experientia* 44: 960–971.

Grobstein P (1989) Organization in the sensorimotor interface: A case study with increased resolution. In Ewert J-P, Arbib MA (eds), *Viscomotor Coordination: Amphibians, Comparisons, Models, Robots.* New York, Plenum, in press.

Grobstein P, Comer C, Hollday M, Archer S (1978) A crossed isthmotectal projection in *Rana pipiens* and its involvement in the ipsilateral visuotectal projection. *Brain Res* 156: 117–123.

Grobstein P, Comer C, Kostyk SK (1983) Frog prey capture behavior: Between sensory maps and directed motor output. In Ewert J-P, Capranica RR, Ingle D (eds), *Advances in Vertebrate Neuroethology.* Plenum, New York.

Holmes G (1945) The organization of the visual cortex in man. (Ferrier Lecture). *Proc Roy Soc Lond* B 132: 348–361.

Humphreys GW, Riddoch MJ (1987) On telling your fruit from your vegetables: A consideration of category-specific deficits after brain damage. *TINS* 10: 145–148.

Ingle D (1983) Brain mechanisms of visual localization by frogs and toads. In Ewert J-P, Capranica RR, Ingle D (eds), *Advances in Vertebrate Neuroethology.* Plenum, New York.

James, W (1980) *Principles of Psychology.* London, Macmillan.

Jeannerod M. (1985) *The Brain Machine.* Cambridge, Harvard University Press.

Kennedy D (1971) Nerve cells and behavior. *Amer Sci* 59: 36–42.

Kolb B, Whishaw IQ (1980) *Fundamentals of Human Neuropsychology.* San Francisco. W H Freeman.

Kostyk SK, Grobstein P (1982) Visual orienting deficits in frogs with various unilateral lesions. *Behav Brain Res* 6: 379–388.

Kostyk SK, Grobstein P (1987a) Neuronal organization underlying visually elicited prey orienting in the frog: I. Effects of various unilateral lesions. *Neuroscience* 21: 41–55.

Kostyk SK, Grobstein P (1987b) Neuronal organization underlying visually elicited prey orienting in the frog: III Evidence for the involvement of an uncrossed descending tectofugal pathway. *Neuroscience* 21: 83–96.

Lashley KS (1950) In search of the engram. *Symp Soc Exp Biol* 4: 478–505.

Lashley KS (1951) The problem of serial order in behavior. Jeffries L (ed), *Cerebral Mechanisms in Behavior.* Wiley, New York.

Loeb GE (1987) Hard lessons in motor control from the mammalian spinal cord. *TINS* 10: 108–113.

Luria AR (1980) *Higher Cortical Functions in Man.* Basic Books, New York.

Marr D (1982) *Vision.* WH Freeman, San Franciso.

Merzenich MM, Jenkins WM, Middlebrooks JC (1984) Observations and hypotheses on special organizational features of the central auditory system. In Edelman GM, Gall WE, Cowan WM (eds) *Dynamic Aspects of Neocortical Function.* New York, Wiley.

Mountcastle, V (1974) (ed) *Medical Physiology.* St. Louis, Mosby.

Mpitsos GJ, Cohan CS (1986) Convergence in a distributed nervous system: parallel processing and self-organization. *J Neurobiol* 17: 517–545.

Oppenheimer JM (1977) Studies of brain asymmetry: Historical perspective. *Ann NY Acad Sci* 299: 4–17.

Pellionisz A (1983) Brain theory: Connecting neurobiology to robotics. *J Theoret Biol* 2: 185–211.

Rosner BS (1974) Recovery of function and localization of function in historical perspective. In Stein DG, Rosen JJ, Butters N (eds), *Plasticity and Recovery of Function in the Central Nervous System*. New York, Academic.

Selverston AI (1980) Are central pattern generators understandable? *Behav Brain Sci* 3: 535–572.

Sherman SM (1974) Visual fields of cats with cortical and tectal lesions. *Science* 185: 355–357.

Sherman SM (1977) The effect of superior colliculus lesions upon the visual fields of cats with cortical ablations. *J Comp Neurol* 172: 211–230.

Sperry RW (1974) Lateral specialization in the surgically separated hemispheres. In Schmitt FO, Worden FG (eds), *The Neurosciences: Third Study Program*. Cambridge, MIT Press.

Sprague JM (1966) Interaction of cortex and superior colliculus in mediation of visually guided behavior in the cat. *Science* 153: 1544–1547.

Stein PSG (1976) Mechanisms of interlimb phase control. In Herman RM, Grillner S, Stein PSG, Stuart DG (eds), *Neural Control of Locomotion*. New York, Plenum.

Teuber H-L, Battersby WS, Bender MB (1960) *Visual Field Defects after Penetrating Missile Wounds of the Brain*. Cambridge, Harvard University Press.

Weiskrantz L (1986) *Blindsight*. New York, Oxford University Press.

Wilson DM (1966) Central nervous mechanisms for the generation of rhythmic behavior in arthropods. In *Nervous and Hormonal Mechanisms of Integration*. New York, Academic.