Theoretical Antecedents

Perception cannot work by extracting production invariants if production works by using perceptual invariants. If there are perceptual invariants for use in production, then they should be used in perception. This buck-passing between production and perception has only been possible because few theorists have attempted to solve the problems of both areas within the same general theory.

(Howell & Harvey, 1983, p. 203)

This chapter describes some conceptual antecedents to the node structure theory of perception and action that I develop in the remainder of the book. Conceptual antecedents to my theory stretch back to Plato, but I mention only a sample of these antecedents here, and this chapter can be viewed as a summary of my personal sources of inspiration rather than as an authoritative historical review (see also Rumelhart, McClelland, & the PDP Research Group, 1986). I begin with two general philosophical traditions that have had longstanding and profound effects on virtually all psychological and physiological thought. I then illustrate how these traditions have influenced three current theories of the relation between perception and action in general and speech perception and production in particular. I next outline a theoretical alternative to these philosophical traditions that Lashley (1951) pointed out and that constitutes a major theme of the present book. Finally, I spell out some more recent and detailed conceptual antecedents to the theory that I go on to develop in the remainder of the book.

Philosophical Antecedents

The relation between perception and action has been debated since the time of Descartes, and two general philosophical views have prevailed in this continuing debate. One is that action is subordinate to and less important than perception, and the other is that perception and action constitute separate domains of inquiry.

The Subordination-of-Action Tradition

Many philosophers have viewed action as functionally, temporally, and evaluatively subordinate to perception; functionally subordinate because they consi-
dered perception the sole means by which knowledge is acquired (empiricism), temporally subordinate because they considered perception a necessary precursor to action (paleobehaviorism), and evaluatively subordinate because they viewed the contemplative life as superior to a life of action (see Plato).

**Effects of Philosophical Subordination**

Effects of the conceptual subordination of action seem predictable in retrospect. The topic of perception has attracted a great deal of attention, whereas the topic of action has been relatively neglected (e.g., Gentner, 1985). Of course, psychologists often give another reason for choosing to study perception rather than action, namely that perception is methodologically easier to study. For example, Fodor, Bever, and Garrett (1974) note that psycholinguistic research has, until very recently, concentrated almost exclusively on perception, rather than on production, or on the relation between perception and production, and they attribute the neglect of production to methodological difficulties.

However, the methodological difficulty hypothesis is clearly incomplete or inadequate as a general explanation of the relative disinterest in action. Recent studies of action (e.g., Sternberg, Monsell, Knoll, & Wright, 1978) are as well controlled as any in perception. Physiology provides another problem for the methodological difficulty hypothesis. Physiologists can trigger actions electrophysiologically and thereby overcome the hypothesized methodological problems, but they too have studied perception more often than production. The methodological difficulty hypothesis also has difficulty explaining reversals in the general trend. For example, whereas perception has received more attention than production in the field at large, the opposite is true in the case of speech errors (Chapter 6). Naturally occurring misperceptions, or slips of the ear, have been collected and studied much less often than naturally occurring misproductions, or slips of the tongue (Fromkin, 1980). Finally, the whole idea that methodological ease represents a viable reason for examining or not examining a general topic area such as perception versus action seems open to question.

**An Assessment of Philosophical Subordination**

Contrary to the long-held philosophical subordination position, everyday perception and action interact with and support one another, and neither can be considered functionally or temporally subordinate to the other. The main function of perceptual and cognitive systems is to guide purposeful actions and to adjust ongoing actions to the situation at hand. As Allport (in press, p. 2) points out, "Perceptual systems have evolved in all species of animals solely as a means of guiding and controlling action, either present or future." Perceptual systems aren't primarily designed to describe and classify the environment in answer to a question such as "What is out there?" but to address the more general questions, "What does it signify for me? What must I do about what's out there?" (after D. M. MacKay, 1984). In short, the nature of the information required for the guidance of actions ultimately determines how perceptual and cognitive systems structure...
the sensory and intellectual environment. Functionally, perception is as subordinate to action as action is to perception.

Temporally also, perception sometimes precedes action, and action sometimes precedes perception. When pricked with a pin, for example, we perceive the pain only after withdrawing the finger; the awareness of pain follows rather than precedes the behavior (James, 1890). And sometimes action proceeds in the absence of perceptual awareness. For example, we continually and automatically use visual cues to orient ourselves in space, stand erect, and perform actions such as walking, even though we never perceive or become aware of these visual cues (D. N. Lee & Lishman, 1974). Usually, however, action and perception take place at the same time. When making saccades in the visual field, for example, perception and action are so intimately intertwined that temporal priority or subordination is impossible to assign.

Even from an evolutionary perspective, perception cannot take precedence over action; systems for perception and action are in general so intimately interrelated as to require mutual adaptation. Consider speech perception-production for example. The capacities for perceiving and producing speech could only have evolved simultaneously. If a mutation suddenly enabled a group of humans to understand language, their chances of surviving to transmit the mutation would only improve if a second group of humans had a mutation that enabled them to speak (see Geschwind, 1983). Moreover, the mutation that enabled this second group of humans to speak would only improve their chances of survival if they had a language to speak and someone to understand them when they spoke. Speech perception could not have evolved before speech production, and vice versa.

The Segregation of Perception and Action

The second major tradition in the history of philosophical ideas relating perception and action is that perception and action constitute separate systems and domains of inquiry. Beginning with Descartes, the afferent processes that mediate perception of the external world (the mind) have been considered separate from the efferent processes that mediate action in the external world (the will). These supposedly separate systems have also been assigned different functions. Perceptual systems are supposed to register and construct a meaning for sensory events, whereas motor systems are supposed to write and execute motor commands. As Turvey (1977) points out, perception and action have virtually no contact with one another in this traditional dichotomy; how a perceptual system perceives neither influences nor is influenced by how the motor system uses perception.

Effects of Philosophical Segregation

Philosophical segregation of action and perception has also had predictable effects. Two separate research areas, with little or no interaction between them,
have developed in parallel, one set specializing in afferent processes, the other in efferent processes. Not just in psychology, but in the many other disciplines interested in perception and action (see the Preface), theories of action have been constructed without reference to perception, and theories of perception have been constructed without reference to action. With some notable exceptions such as the perceptual learning theory of Held and Hein (1963), virtually no theories have attempted to solve the problems of both perception and action at the same time (Howell & Harvey, 1983), and even Held and Hein's (1963) theory assumes completely separate units for perception versus action.

Mirroring the segregation tradition in philosophy, psychologists have not only chosen to study perception more often than action, they by and large have attempted to study perception in the absence of action. Perceptual experiments are characteristically designed to exclude the possibility of action, and this perception-without-action approach has almost certainly influenced the nature of the perceptual systems examined. For example, perceptual experiments have focused mainly on vision and typically attempt to eliminate action by using presentation times that are so brief as to prevent the possibility of eye movements. Touch, on the other hand, has received relatively little attention, perhaps because touch confounds the traditional dichotomy between perception and action: Movement of the hand and tactile perception of an object are contemporaneous and cannot be separately examined or factored out in everyday tactile perception. Even though touch is phylogenetically older and more basic than either vision or speech (von Bekesy, 1967), the perception-without-action approach must avoid touch as impossible to study in vacuo.

ASSESSMENT OF PHILOSOPHICAL SEGREGATION

Needless to say, the perception-without-action approach is by definition unsuited for studying relations between perception and action. However, recent developments in many disciplines have contradicted the long-held view that perception and action are completely separate and call for a new approach to the whole topic. For example, recent neurophysiological research has made it increasingly obvious that the traditional distinction between afferent versus efferent processes in the cortex can no longer be usefully maintained. For example, Ojemann (1983) discovered cortical sites where electrical stimulation interferes with both the perception and production of everyday actions. The sensory and motor areas are inseparable in these and other studies, as if some of the units responsible for afference and efference in the cortex are identical.

Theories Incorporating the Segregation Assumption

To illustrate how the philosophical segregation of perception and action has influenced psychological theories, I examine three well-known theories: the classical theory, the motor theory of speech perception, and feedback control theory.
Theories Incorporating the Segregation Assumption

In its application to speech perception–production, like other theories of perception–production, all three theories explicitly attempt to relate perception and action via separate rather than shared components for perception versus production. Beyond this, the three theories are remarkably different. They postulate different mechanisms, and they deal with different perception–production issues. Their only other shared characteristic is a state of crisis; all three have encountered fundamental phenomena that contradict their basic assumptions. However, I do not attempt to systematically describe or present criticisms of these theories. Nor do I compare these theories with my own. Only after I have developed the relevant aspects of my own theory do I compare it with other more recently published theories of perception (e.g., McClelland & Elman, 1986) and of action (e.g., Norman & Rumelhart, 1983).

The Classical Theory

The classical theory of the relation between speech perception and production holds that the two systems employ separate components at every level of processing (see Straight, 1980). Broca and Wernicke pioneered this theory (D. G. MacKay et al., 1987): They argued from studies of left hemisphere brain injuries that production is localized in one area of the brain and perception in another, interconnected but separate area. However, recent studies using a variety of new and more sophisticated techniques suggest that the picture is more complicated. Expressive and receptive deficits are usually commensurate in extent. For example, with appropriate controls for lesion size, aphasics with severely impaired production also display severely impaired comprehension and vice versa (Mateer, 1983). Moreover, deficits in perception and production are usually similar in nature. Production deficits tend to be more obvious than perceptual deficits in everyday life, because aphasics can simulate comprehension using nonlinguistic cues. However, when given sophisticated tests of comprehension with appropriate controls for semantic and pragmatic cues, Broca's aphasics display comprehension deficits that parallel their more readily observed production deficits. As W. E. Cooper and Zurif (1983, p. 228) point out, "Recent studies are in agreement in concluding that, to the extent that Broca's aphasics show relatively intact comprehension, it is largely based on their ability to utilize semantic and pragmatic cues independent of sentence structure."

The comprehension deficits of Wernicke's aphasics are likewise matched by production deficits, usually involving sentential rather than phonological units (Blumstein, 1973). Difficulties with word order are a typical problem, word salads representing the most extreme case. Although the sentential intonation of Wernicke's aphasics often sounds normal, their speech typically lacks content, contains neologicistic or nonsense elements, and shows errors in sound and meaning. Contrary to the classical theory, Wernicke's aphasics are agrammatic in production as well as comprehension (W. E. Cooper & Zurif, 1983).

Ojemann's (1983) recent findings using cortical stimulation techniques present another problem for the classical theory. Ojemann (1983) discovered sites where
electrical stimulation interferes simultaneously with the mimicking of orofacial gestures and with the perceptual identification of phonemes, as if the same units played a role in both perception and production. Finally, brain scan and cerebral blood flow studies indicate that Broca's area (which under the classical theory only becomes active during production) also becomes active during comprehension (Lassen & Larsen, 1980). All of these findings are less consistent with the classical theory than with Lashley's (1951) hypothesis that common components underlie the perception and production of speech (see also Colthart & Funnell, 1987; Meyer & Gordon, 1983).

The Early Motor Theory of Speech Perception

The early motor theory of speech perception recognized the importance of interaction between the systems for perceiving and producing speech. Motor units that are (necessarily) distinct from the perceptual units can come to the aid of the perceptual units under the motor theory. That is, speech perception and production employ separate components, but at least some speech sounds are perceived with the help of the components that are used for producing them (Liberman, Cooper, Harris, & MacNeilage, 1962; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

As Howell and Harvey (1983, p. 215) point out, “Motor theory attempted to explain something about which very little was known (i.e., speech perception) in terms of something else about which even less was known (i.e., speech production). The problems associated with it are legion.” One of the currently unresolved problems concerns the logical basis of the theory. In order for a pattern of acoustic energy to call up its appropriate production components, a full-fledged perceptual analysis is necessary (e.g., Pick & Saltzman, 1978). This brings the basis for the theory into question because a full-fledged perceptual analysis prior to motor consultation means that perceptual components can accomplish speech recognition without help from the motor components.

Feedback Control Theory

Feedback control theory is in some sense a converse of the motor theory. The perceptual mechanisms come to the aid of the production mechanisms under feedback control theory because perceptual feedback plays a direct and necessary role in producing ongoing speech or action (Adams, 1976; Schmidt, 1982). For example, under feedback control theory, auditory feedback from the pr in the word production could function to trigger production of the o and so on for the remainder of the word.

As expected under this and other theories, an intact auditory system is necessary for the acquisition of normal speech production. However, once speech has been acquired, eliminating or distorting auditory feedback has little effect on the ability to produce intelligible speech (Siegel & Pick, 1974), which suggests that sensory feedback may be unnecessary for well-practiced speech production.
Articulatory disruption does occur when auditory feedback is amplified and delayed (B. S. Lee, 1950), but even here, feedback control theory fails to fit the detailed nature of this phenomenon (reviewed in Chapter 10).

Lashley's Alternative: Common Components

Not all theoretical thinking in psychology has adopted the assumption that components for perception and action are completely separate or unshared. In particular, Lashley (1951, p. 186) proposed that speech comprehension and production make use of common components and mechanisms because "the processes of comprehension and production of speech have too much in common to depend on wholly different mechanisms."

Like Lashley, I am especially concerned in the present volume with the concept of shared components for perceiving and producing speech. After reviewing the available evidence in Chapter 2, I conclude that perception and production share some components but not others. One system of unshared components represents patterns of sensory input for perception, and another system represents patterns of muscle movement for production, while the systems of shared perception–production components represent phonological units such as segments and syllables and sentential units such as words and phrases.

However, the concept of integrated or shared perception–production mechanisms is not restricted to speech. For example, Darian-Smith, Sugitani, and Heywood (1982) discovered cells in the somatosensory cortex that respond both to finger movement and to sensory properties of a textured surface. Indeed, Pribram (1971) viewed the so-called motor cortex as a system for somatosensory regulation as well as for action, arguing that the motor representation of an action must contain a perceptual "image of achievement," because processing of feedback is necessary to ensure that the action has been executed as intended. Supporting this view, Pribram (1971) and Kornhuber (1974) reviewed evidence indicating that cells in the motor cortex are responsive to cutaneous and somatosensory stimuli from the body part moved.

Self-perceptions of actions induced by cortical stimulation are consistent with these observations. For example, when motor cortex stimulation results in arm movement, subjects never report an introspective sequence beginning with an urge to move the arm, followed by arm movement, and ending with perception of arm movement. A cortically induced movement seems unwilled, happens by itself, and is perceived and performed simultaneously (Bridgeman, 1986).

Puzzling Asymmetries Between Perception and Action

Asymmetries between perception versus production processes represent the main theoretical puzzle or challenge facing the idea of shared perception–production units and may explain in part why virtually no psychological theories have taken up Lashley's (1951) suggestion. An example of asymmetry is the fact
1. Theoretical Antecedents

(discussed in greater detail in Chapter 6) that speech perception can proceed much more quickly than speech production. Computer-compressed speech remains perceptually intelligible at five to seven times the rate that people can produce speech of comparable intelligibility (Foulke & Sticht, 1969). This rate asymmetry cannot be completely explained in terms of the muscular or biomechanical factors involved in speech production but reflects a central and inherent processing difference that must be explained in theories of speech perception-production (Chapter 6).

What processing differences could enable perception to proceed so much faster than production? The node structure theory postulates several fundamental processing differences (summarized in Chapter 6) that together explain not only the rate asymmetry but many other asymmetries between perception and action, such as perceptual differences between self-produced versus other-produced feedback, differential effects of practice on perception versus production, and asymmetries in the nature of the errors that occur in perception versus production.

Antecedents to the Processing Characteristics of the Node Structure Theory

The basic components of the node structure theory are nodes, a psychological term that dates back at least to Collins and Quillian (1969). However, my use of the term nodes resembles Wickelgren's (1979) and McClelland and Rumelhart's (1981) rather than Collins and Quillian's (1969), Estes' (1972), or Anderson and Bower's (1973). Nodes in the latter writings refer to intersections in a parsing tree and represent descriptive rather than theoretical terms, whereas nodes in the node structure theory are theoretical constructs, processing units that share the same structural characteristics and dynamic or processing capabilities and respond in the same way to basic variables such as practice (repeated activation). Here I discuss the dynamic characteristics of nodes, which go well beyond the concept of an intersection and have other historical antecedents of their own. The remainder of the book then examines how these dynamic characteristics become implemented during everyday perception and action.

Dynamic Characteristics of Nodes

Nodes have five dynamic properties that are relevant to all aspects of the organization of perception and action: activation, priming, satiation, self-inhibition, and linkage strength. Each of these dynamic properties, taken by itself as in the discussion below, is remarkably simple, but interactions between dynamic properties can be quite complex. Each property influences the others in complex ways that depend on the current state of the node and on its history of activity over the course of a lifetime. Also, perception and action use these dynamic processes differently in the theory, and these processing differences contribute to
already observed perception–production differences, such as the maximal rate asymmetry, and predict new asymmetries for future test. Illustrating the systematics in how the dynamic properties of nodes interact will occupy much of the remainder of the book.

**NODE ACTIVATION**

I use the term *activation* as short for *node activation*, a process necessarily for both perception and action in the theory. My use of the term activation mirrors that of Lashley (1951) but differs from many other current uses of the same term in the cognitive and brain sciences. I attempt to avoid terminological confusion at the outset by comparing my usage with these other concepts of activation.

Like neural activation, node activation is all-or-none and self-sustained. Activation lasts for a specifiable period of time, independent of whether the sources that led originally to activation continue to provide input. However, node activation can, and in the case of mental or perception–production nodes, invariably does, involve activation of more than one neuron. Neurons and nodes also differ greatly in how long they remain activated and in their recovery time following activation. For example, neurons require at most a few milliseconds to recover from activation, whereas nodes require anywhere from a few milliseconds to hundreds of milliseconds (Chapter 8).

Node activation also differs from the concept of spreading activation in propositional network theories such as Anderson (1983). Node activation never “spreads,” and its intensity never changes. Unlike spreading activation, node activation remains constant with “distance,” fatigue, and the number of other nodes that an activated node is connected to. Moreover, node activation is sequential and nonautomatic in nature. A special activating mechanism must become engaged to determine when and in what order different nodes in my theory become activated. By way of illustration, the numbers in Figure 1.1 represent the typical order in which the adjacent nodes become activated during production.

During its period of self-sustained activation, a node simultaneously primes all nodes connected to it, and as we will see, priming is necessary in order to activate a node. Another characteristic distinguishing node activation from spreading activation is the occurrence of self-inhibition (discussed in a following section), a brief period of reduced excitability that follows node activation.

**NODE PRIMING**

My use of the term *priming* also dates back to Lashley (1951). Priming refers to a transmission across a connection that increases subthreshold activity and prepares the connected node for possible activation. Because nodes must become primed in order to become activated, priming is a necessary precursor to all perception and action. An activated node simultaneously primes all nodes connected directly to it, and nodes that are “once removed” from an activated node also receive priming but to a lesser extent. Thus, priming falls off sharply in degree with distance from the source. An activated node primes its connected nodes
most strongly (first-order priming), while a node receiving first-order priming primes its connected nodes less strongly (second-order priming). Third-order priming arising from the activation of a single node is negligible in degree, and unless it summates with priming from other sources, third-order priming can be ignored in theories of production. Thus, priming spreads, but only to a limited degree, and unlike propositional network theories, activation never spreads at all.

Priming summates across all simultaneously active connections (spatial summation), and priming accumulates during the time that any given connection remains active (temporal summation). Consider, for example, the temporal summation of top-down priming across the connections illustrated in Figure 1.1. Node 1 becomes activated first and simultaneously primes nodes 2 and 5 (Figure 1.1). However, node 5 cannot become activated until nodes 2, 3, and 4 have been activated. Priming of node 5 therefore continues to summate during the time that nodes 2, 3, and 4 are being activated. The anticipatory nature of this accumulating priming facilitates the eventual activation of node 5 and all other “right-branching” nodes at every level in such a hierarchy. Because virtually all top-down connections are divergent (one-to-many), top-down anticipatory effects are universal in the theory (D. G. MacKay, 1982). As we will see, however, these anticipatory effects incur a built-in cost. Temporal summation of priming increases the probability of “anticipatory errors,” where an about-to-be-produced component is produced before its time, the most common class of errors at either the phonological or sentential levels of speech production (e.g., Cohen, 1967).
Primming only summates to some subthreshold asymptotic level (Figure 1.2), and cannot by itself cause a connected node to become activated. As a consequence, primming never results in behavior when the lowest level muscle movement nodes representing an action become primed.

Unlike activation, primming cannot be self-sustained and begins to decay as soon as the activity of a connected node stops (Figure 1.2). Also unlike activation, primming is not followed by a period of self-inhibition and recovery. Finally, primming is untimed and order free. No special triggering mechanism is required to determine when and in what order nodes become primed. In this sense, primming is automatic and parallel in nature, whereas activation is controlled and sequential.

Despite these many differences, activation and primming are intimately related in the theory. A minimal degree of primming is necessary for activation. Unless a node achieves this minimum primming level, designated the commitment threshold, its activation mechanism will be unable to activate it. However, achieving the minimal level of primming is insufficient to guarantee activation if and when the activation mechanism is applied. In order to become activated,
a node must also receive more priming than any other node in its domain. Chapter 2 discusses the theoretical and empirical basis for this “most-primed-wins” principle, the means by which all nodes become activated.

**Self-Inhibition**

After nodes representing the components of skilled behavior become activated, they undergo a brief period of self-inhibition, during which their level of priming falls below normal or resting level (Figure 1.2). Following self-inhibition, nodes undergo a recovery cycle consisting of a hyperexcitability phase followed by a return to resting level. During the hyperexcitability phase, or postinhibitory rebound, priming first rises above and then returns to resting level (Figure 1.2). Various sources of evidence bearing on the time characteristics of self-inhibition and the recovery cycle are discussed in Chapter 8.

**Satiation**

Satiation refers to a fatigue process during which a node becomes less responsive to priming. This reduced sensitivity occurs when the node has been activated repeatedly over a prolonged period of time, say 5 to 30 s, and manifests itself in two ways. The duration of self-inhibition following activation becomes extended, and the rebound from self-inhibition falls below normal or resting level. Satiation of course varies in degree, depending on the extent and duration of repeated activation.

**Linkage Strength**

The concept of linkage strength also has a long and distinguished history, dating back at least to Thorndike (1898). Linkage strength represents a relatively long-term characteristic of a connection that has been used to explain a wide range of practice effects in the psychological literature (D. G. MacKay, 1982). Practice, or more specifically, the frequency with which a node has been primed and activated via a particular connection in the past determines linkage strength in the node structure theory. However, linkage strength has a special relation to priming in the theory. Connections with high linkage strength transmit priming more rapidly and provide more priming at asymptote than do connections with low linkage strength. That is, linkage strength influences how fast priming will summate across a connection per unit time (represented by the initial slope of a priming function, such as the one illustrated in Figure 1.2), and linkage strength influences how much priming can be transmitted across the connection before asymptote is reached.

In summary, the dynamic properties of nodes (activation, self-inhibition, priming, satiation, and linkage strength) are closely interrelated. Priming is necessary for activating a node, and the degree of priming is related to the probability of activation in ways discussed in subsequent chapters. Activating a node increases the linkage strength of its connections and causes its connected nodes to become
primed. Linkage strength in turn influences how much and how rapidly priming can be transmitted across a connection. Finally, self-inhibition terminates activation and, during satiation, is itself influenced by activation.

**Special Dynamic Properties**

In addition to the universal dynamic properties discussed previously, some nodes also have special dynamic properties, such as quenching, multiplication of priming, and periodicity. These special dynamic properties differentiate three classes of nodes: content nodes, sequence nodes, and timing nodes. Content, sequence, and timing nodes also differ in how they connect with other nodes and in the functions they perform in perception and action. *Content nodes* represent the form or content components of an action or perception. *Sequence nodes* activate content nodes in some specifiable sequence. And *timing nodes* activate sequence nodes at some specifiable rate.

**Quenching**

Quenching is a special characteristic of content nodes, with conceptual antecedents in Grossberg (1982). Once a content node becomes activated, it quenches or inhibits the sequence node that originally caused it to become activated. The next chapter discusses reasons for including this quenching mechanism in the theory.

**Multiplication of Priming**

Multiplication of priming is another process anticipated in some respects by Grossberg (1982) and represents a special function carried out by sequence nodes. An activated sequence node doesn't simply prime its connected nodes; it multiplies their existing level of priming by some factor per unit time. For example, a sequence node might double the level of priming of a connected content node every 2 ms. This multiplication of priming process provides the basis for the most-primed-wins principle by which all content nodes become activated.

**Periodicity**

Periodicity is a process with a long theoretical history and refers in the node structure theory to an endogenous and inherently rhythmic pattern of activation that is characteristic of timing nodes. Once a timing node becomes engaged, it spontaneously self-activates every (say) 10 ms.