1 Introduction

This book outlines an approach to the development of expressive and communicative behavior in early infancy until the onset of a single word which is rooted in ethology and dynamic action theory. Here the process of expressive and communicative actions, organized as a complex and cooperative system with other elements of the infant's physiology, behavior and social environments, is elucidated. Overall, humans are provided with a finite set of specific behavior patterns, each of which is probably phylogenetically inherited as a primate species. However, the patterns are uniquely organized during ontogeny and a coordinated structure emerges, which eventually leads us to acquire spoken language. A dynamic model is presented where elements can be assembled for the onset of language in the infant in a more fluid, task-specific manner determined equally by the maturational status and experiences of the infant and by the current context of the action.

No doubt, communication is a social phenomenon and the most prominent feature of human speech and language. The complex organization of human societies is mediated by the ability of members to inform one another and is dependent on the exchange of information. Therefore, not surprisingly, many scientists have focused attention on how children acquire language ability.

Although children do not produce linguistically meaningful sounds or signs until they are approximately one year old, the ability to produce them begins to develop in early infancy, and important developments in the production of language occur throughout the first year of life. Unless they are hearing-impaired, infants acquire phonology during their first year. In spoken language, the acquisition of phonology consists of learning to distinguish and produce the sound patterns of the adult language. At birth, the newborn has the ability to distinguish virtually all sounds used in all languages, at least when the sounds are presented in isolation. The newborn produces no speech sounds, however. During the first year of life, speech-like sounds gradually emerge, beginning with vowel-like coos at six to eight weeks of age, followed by some consonant sounds, then followed by true babbling. By the end of the first year, children are typically babbling sequences of syllables that have the intonation contour

of their target languages. Finally, meaningful words are produced; that is, the onset of speech occurs.

The factors that underlie these developments include: physical growth of the vocal apparatus, neurological development, and language experience. Language experience exerts its influence on both the perception and the production of speech sounds. Characteristics of the vocal apparatus that enable us to acquire language, features of neurological development, and features of the manner in which the experience of ambient language influences children's linguistic behavior are all uniquely human, and this uniqueness can only be adequately comprehended when we view the process of early language development from a comparative perspective. Moreover, the predisposition of humans to acquire language is not restricted to a specific modality but rather is somewhat amodal. When humans have difficulty acquiring spoken language, other possibilities can be explored – a further biological predisposition that has phylogenetically evolved exclusively in humans.

A primate behaviorist's view of language acquisition

By comparing human language with the communicative behavior of nonhuman primates, this book will take an ethological perspective in exploring the changes that occur during this earliest stage of language development. Animal societies are equally dependent on the exchange of information. Any organism that lives in complex social groupings must rely on communicating some aspects of its status to others. Such an exchange of information, the process that defines a communication system, implies the existence of a common language or a common set of rules that govern the encoding and decoding of signals in the communication system.

It is tempting to think of animal communication systems as being composed of simple invariant designators or external manifestations of some basic internal states such as hunger, pain or reproductive readiness. For nonhuman primates, however, it is known that, in addition to these states, many other individual and societal factors such as individual identities, kinship, roles, dominance relations and coalitions play an important part in social organization and social behavior. The complexity of many primate societies kindled interest in the communication systems mediating social behavior. For this reason, the objective and quantitative description of vocal communication began earlier in nonhuman primate studies than in studies of human infants.

Carpenter (1934), a pioneering researcher, introduced in his observations of howler monkeys the basic method that is still used – describing vocalizations and the situations in which they were used. Rowell and Hinde (1962) were the first to characterize the vocal repertoire of a monkey, the rhesus macaque, by publishing sound spectrograms. Winter, Ploog and Latta (1966) added a

quantitative dimension to the analysis by measuring acoustic features of the sounds recorded in their colony of squirrel monkeys. Struhasaker (1967) statistically analyzed the vocalizations recorded in his field study of vervet monkeys.

As a primate behaviorist, these early pioneering works influenced my initial interest in language. Consequently, my first exposure to the study of language did not involve human infants, children or even adults. In 1979, I was living in the upper Amazonian basin in Bolivia observing groups of a free-ranging New World primate, Goeldi's monkey. While there, I recorded their vocalizations. During my observations, I found that the animals exhibited two different types of responses when group members encountered a predator and emitted an alarm call. One was to climb down to the ground and to freeze there. The other was to climb up to the highest strata in the canopy and to mob. Different types of alarm calls appeared to be associated with different types of predators and the behavioral responses were assumed to vary with call type. However, the sound spectrographic analyses that I conducted upon returning to Japan showed that the entire sample of alarm calls fell along a graded continuum. Therefore, I chose to focus my doctoral thesis on how Goeldi's monkeys perceive conspecific alarm calls. Using captive animals, I investigated their responses to experimentally produced conspecific natural calls as well as to synthesized versions of them that varied in the acoustic parameters that defined the calls under study. Although natural alarm calls showed considerable individual heterogeneity, playbacks of synthesized versions of these calls that varied in a single acoustic parameter produced gross differences in behavioral responding across a narrow acoustic boundary.

With respect to speech perception in humans, if one creates synthetic speech stimuli representing equal steps along the continuum of a single acoustic parameter (for example, voice-onset-time ranging from simultaneous voicing to increasingly delayed voicing) and plays these stimuli to subjects, subjects report the experience of hearing either of two different sounds (for example, /ba/ or /pa/) rather than a graded series of sounds. That is, they perceptually group several different stimuli as /ba/ and certain other stimuli as /pa/. There is no apparent ambiguity between /ba/ and /pa/. A given stimulus from any point on the continuum is labeled as one or the other phoneme, and the two phonemes are strictly categorized; this phenomenon is known as categorical perception. The findings I obtained on vocal perception in Goeldi's monkeys appear analogous to this categorical perception that humans demonstrate with speech sounds, though at present such a perception is thought to be restricted to speech sounds.

After earning my doctorate, I briefly conducted research in Texas, USA. There, I investigated the perception of conspecific alarm calls in a group of Japanese macaques that had been translocated from the Kyoto area of Japan ten years prior. In my work with Japanese macaques, I employed the same

experimental paradigm as in my previous work with Goeldi's monkeys. I found that Japanese macaques also perceive their conspecific alarm calls categorically, as demonstrated in human speech perception. From my studies, I learned that what is perceived as a single unit of behavior by human observers (i.e., what is heard as a single class of vocalization, in this case) may not actually be perceived as such by members of other species. These findings, together with similar results with other nonhuman primate species (see Snowdon 1982, for review), were rather astonishing because previous researchers attempting to construct vocal repertoires for nonhuman primate species (e.g. Rowell and Hinde, 1962) have noted the complex call structure of animals that was highly variable both between individuals and within the repertoire of a single individual. That is, many calls could not be easily categorized into discrete classes but rather call structures seemed to intergrade with one another. Researchers have assumed that in many cases these intergradations corresponded to hypothetical underlying motivational continua, thus the intergrading call structure was said to map a continuous motivational system. Despite this sort of variability and complexity, findings like my own suggest that we must be very cautious about how we define units of behavior in nonhuman primates. Based on such reflection, thereafter, primatologists working with vocal communication started to seek new methodologies that could reconcile the continuous variability in calls with the discrete messages they appear to carry. In addition, they successfully expanded the notion of vocal communication in traditional ethology. In so doing, they sought to elucidate the evolutionary continuity between nonhuman primate vocalization and human language.

Implications and limits of the traditional ethological approach to communication

The term "ethology" refers to the biological study of behavior (Tinbergen, 1951). It has been claimed that the discipline of ethology offers a unique integration of a unifying theory, evolutionary biology, with a methodological heritage, naturalistic observation (Blurton-Jones, 1972; Charlesworth, 1980). The operational translation of the evolutionary perspective on to behavior was provided by an early pioneer of ethology, Nicholas Tinbergen. Tinbergen (1951) defined ethology as follows:

> the science [of ethology] is characterized by an observable phenomenon (behavior, or movement), and by a type of approach, a method of study (the biological method). The first means that the starting point of our work has been and remains inductive, for which description of observable phenomena is required. The biological method is characterized by the general scientific method, and in addition by the kind of questions we ask, which are the same throughout Biology and some of which are peculiar to it. (1951: 411)

The modern synthetic theory of evolution provides an integrative framework for many disciplines and content areas. Naturalistic observation provides not only essential descriptive data but it also serves as an invaluable source of ecologically valid hypotheses. Current ethology does not stress biological determinism but rather a multilevel perspective that can expand and enrich our understanding of development. Tinbergen argued that the question, "Why does this animal behave in this way?" included four different questions in the "why." The first question asks why the animal performed a particular behavior now, the question of immediate causal control of the behavior. The second question asks how the animal grew to respond in that particular way, the question of individual development. The third question asks why this kind of animal does this particular behavior, the question of survival value or function of the behavior. Finally, there is the question of why this group of animals came to solve this problem of survival in this way, the question of evolutionary origins of the behavior.

Until the mid-1980s, virtually all investigators interested in the vocal communication systems of nonhuman primates were concerned with the problem of human language in terms of these four questions. Those engaging in research with nonhuman primates looked for clues to illuminate the evolutionary background and biological heritage of human language. These kinds of clues, hints of the rules by which socially important information is encoded into and decoded from speech sounds, are especially relevant to hypotheses on the origins of human language since there are no fossil records available and one has to rely on comparative studies alone. The uses of vocalizations and their relationship to social behavior may be investigated when both the auditory and social parameters of behavior are available. In fact, in many nonhuman primates, certain features of the social situations in which the sounds are emitted are accessible to the investigator.

The approach to language that I adopt in this book might surprise those who have little knowledge about recent advances in primatology with respect to vocal communication. For example, linguists and developmental psychologists who regard language as a capability beyond the reach of animal research subjects might conclude that primate vocal communication falls outside their own purview as investigators and scholars. Such reactions would not be unexpected given that mainstream modern linguistics has been more concerned with theories of grammar than social communication and ecologically valid models of language use. Further, language has also been defined in very abstract terms and treated by many linguists as though it were synonymous with generative morphology and syntax.

By considering the general characteristics of vocal systems and how they are used, a number of primatologists interested in communicative behavior have recently revived the traditional ethological paradigm in order to place the

interspecies comparison of vocal sounds in perspective for nonhuman primates. The conceptual framework for this book is inspired by the theories and methods of this recently expanded ethology as well as by current knowledge about vocal communication in nonhuman primates. The arguments raised and the paradigms developed in recent research also contribute to our understanding of the nature of linguistic capacity and are particularly indispensable to understanding how preverbal human infants acquire language. However, before I explore arguments surrounding language development in human infants, I will outline recent advancements in research on nonhuman primate vocal communication. A focus on such research will help show why evolutionary and comparative perspectives as formulated in the discipline of ethology are crucial to guide a program of developmental research on humans in general. Indeed, this is particularly important in that recent trends in developmental psycholinguistics research cast nonhuman primates in a more interesting light than ever before.

It is now recognized that language, whether spoken or signed, rests on several different types of motor and phonetic learning systems and a range of potentially contributory precursive behaviors (Bullowa, 1979; Papoušek, Jürgens and Papoušek, 1992; Oller, 2000; Speidel and Nelson, 1989). Hence, it is now deemed legitimate to investigate infants' cognitive and neural development as well as their social perceptual experiences in the quest for understanding how and why they begin to speak. Such an approach is also a theoretical necessity. That is, if infants engage in behaviors that facilitate language before they possess the cognitive capability to fully appreciate its existence, then their behaviors must be motivated by one or more non-linguistic factors (Locke and Snow, 1997). Merely owning the genes of a species known to possess the capacity for language would be insufficient. Linguists have argued that language requires specialized mental mechanisms that are encapsulated or dissociated from other, more generalized processing systems. However, linguists have not yet presented actual evidence for this. I propose that an ethological approach to language development provides one possibility for a breakthrough on this issue.

Discrepancy between ethologists' traditional view and linguists' view of human speech

In his formulation, Tinbergen aptly recognized that a full understanding of behavior includes both proximate and distal "causes" and that one must always view individual animals within the ecological context of the species. In sharing this view, my purpose in this book is in part to illustrate how Tinbergen's formulation can be used to direct research on a class of common, but puzzling infant behavior: language acquisition. That a combination of evolutionary

biology and naturalistic observation potentially has much to offer our understanding of human behavior has been pointed out a number of times over the past few decades. However, Tinbergen's formulation has only been successfully extended to human behavior, more specifically, human language, in just a few investigations. As partial explanation for this, Tinbergen also cautioned that one should not confuse questions asked at one level with those asked at another. For example, Blurton-Jones (1972) argued that the persistence of unproductive nature–nurture arguments in behavioral research is a consequence of the confusion between issues of development and those of adaptation and evolution. More importantly with respect to communicative behavior, it must be acknowledged that ethologists have not understood how linguists distinguish human language from nonhuman communicative behavior on the one hand, and that linguists have not understood the significance of the ethologists' view of language on the other.

Traditional ethology conceived of animal communication as genetically fixed, developmentally immutable, stereotyped activity. Within the communicative repertoire of a species there were thought to be only a relatively small number of invariant signals (Moynihan, 1970) that were used in an equally small number of motivational or contextual situations (Smith, 1977). Although the critical importance of context in the interpretation of signals has been recognized for many years, the prevailing view that has been provided of communication in nonhuman animals has been of a restricted signal repertoire and a restricted set of communicative referents.

According to the traditional ethological view, which assumes discontinuity between human and animal communication, human communication is not stereotyped and is considerably modifiable during development. Human communication employs a signal repertoire of enormous size compared with the repertoires of nonhuman species. Human communication has signal invariants that are easily perceived by human recipients even though it is often difficult for humans to discern the physical structure of signals. If one ascribes to this view, one cannot analyze human communication from an ethological perspective. Earlier studies of sounds produced by nonhuman animals (other than primates) also confirmed that these sounds could be regarded as a sort of fixed action pattern. Before sound spectrum analysis became possible in the 1950s, all sounds were identified by labels that were often idiosyncratic to the person who used them. With the new method, different individuals were now able to agree on the pattern of a signal based on its objective and permanent representation. Pioneering sound spectrographic analyses revealed that many of the vocalizations recorded from a number of bird species could be easily discriminated from one another. However, as noted by Rowell and Hinde (1962), nonhuman primate vocalizations frequently appeared to intergrade with one another and hence were not clearly classifiable into discrete categories.

Therefore, in the ethological view, nonhuman primate vocalizations should be classified into the category of human communication because of their signal feature of forming a graded continuum. However, the ethologists were so naive regarding linguistics in general that they failed to appreciate that the human system does not necessarily use continuous units exclusively. On the contrary, although language employs continuous parameters whereby small changes in acoustic value result in corresponding changes in transmission value (e.g., as one raises one's voice gradually, one may sound increasingly angry or upset), such continuous variations merely correspond to "paralinguistic" signaling. They are not regarded by linguists as playing a role in differentiating lexical items. Linguists concluded that while nonhuman primate vocal communication systems appear in some cases to rely heavily on signal dimensions that vary continuously for communicative value, human vocal communication systems maintain a fundamental distinction between dimensions that are manipulated continuously for paralinguistic effect and segmental features. Moreover, in linguistics, the latter are treated as phonetic units and are interpreted categorically in terms of their lexical effect.

A typical expression of this sort of linguistic view of nonhuman primate vocalizations and human language is Hockett's (1960) characterization of human language, as a communication system, in terms of "design features" (e.g., "discreteness" and "duality of patterning"). According to Hockett, the human system possesses discreteness in that the alphabet-level (segmental phonetic) units have categorical values. That is, a change in the acoustic characteristics of one sound segment (say the b in "bay") is regarded as irrelevant from the standpoint of transmission value (meaning) unless it precedes a shift to a new meaning category (say "pay"). Human language usually includes lexicons of thousands of words constructed from such discrete alphabetic/phonetic units. Nonhuman vocal communication systems often include an inventory of discrete calls or call types (e.g., one for threat, one for affinity, one for alarm). However, their categorical lexicon is usually small in number of meaningful units by comparison with human languages, and importantly, as already noted, it is usually characterized by stereotypy.

The power of the human system to create an extensive lexicon lies in its dependency on the duality of patterning referred to by Hockett. According to Hockett, duality of patterning concerns individual alphabetic units of the human phonetic/phonemic system that are independent of meaning; duality of patterning refers to the fact that these units can be recombined and reordered to construct different units of meaning. Thus the words act, cat and tac(k) all share the same phonemic units while lexically they are entirely distinct.

It is important to emphasize the "recombinability/reorderbility" characteristic implied by this duality because recombinability enables a small number of

phonemic units to be utilized to create an enormous lexicon, by merely stringing the phonemic units in unique patterns. With respect to potential recombinability, studies of nonhuman primate vocal systems appear to show either that no restructuring is possible or that changes are far more limited than those that can occur in human speech. A system that has no recombinability is restricted to a lexical inventory size, which can be no greater than the number of discrete units in the system.

Thus, the use of continuous variations of sounds for communicative purposes that has been recognized in nonhuman species is indeed shared by humans, but in humans continuous variation is only used as a paralinguistic component of vocal communication and not as a component of language itself. Humans also apparently differ from nonhuman primates in making greater use of the categorical features of sound in their vocal communication. Linguists have assumed that through the acquisition of such distinct means, humans exclusively are equipped to produce and use language. The evolution of language is thought to have occurred some time after the emergence of vocal communication like that found in living nonhuman primates, for instance after the acquisition of a unique vocal apparatus as bipedal walkers. In order to produce sounds with the features needed for language, sounds generated by the air stream must be morphologically chopped by vibrating vocal folds.

Methodological characteristics of ethology in investigating nonhuman primate vocalizations

Hockett initially proposed his model in order to criticize naive comparisons between nonhuman sounds and human language. However, having rejected the position of traditional ethologists, one might revisit the original question: how are nonhuman sounds similar to or different from the sounds of human language? Hockett's model provides a framework for discussing only how the sounds "function" (similarly or differently in humans and nonhuman species) but it does not really address the issue of the relationship between human and nonhuman sounds per se. In order to investigate how preverbal infants come to produce sounds that characterize human language, a purely acoustic description of preverbal infant vocalizations could still be meaningful. In this regard, findings obtained from comparisons between the vocal sounds of humans and nonhuman primates could offer an important perspective.

Further, mostly owing to our ever-developing knowledge of human speech perception, the distinction between discrete and continuous vocalizations has blurred recently. Knowledge concerning human speech perception came first from findings on categorical perception, a topic in which I was interested in my doctoral work. Namely, several of our speech sounds appear to form a continuous distribution when examined spectrographically and yet we rarely

have difficulty distinguishing the category into which a particular sound falls. Findings such as these make it difficult to apply the graded-discrete distinction between the signals of primates (humans included) versus the signals of other animals as was done in the earliest nonhuman primate vocalization studies. Whether a repertoire appears large or small depends on how one characterizes signals and how one deals with graded signals. Along with improvements in the detection of signals, early estimates of repertoire size have been altered; while obviously valuable in itself, this has made it even more difficult to draw any conclusions about repertoire size.

In response to the oversimplified dichotomy between animal and human communication, primate behaviorists have sought methods to identify more precisely each call type within a vocal repertoire. As a result, advancements have been made in the techniques used to analyze vocalizations. These advancements fall primarily into three domains that I will discuss presently: (1) contextual analysis, (2) sorting techniques, and (3) playback techniques.

Contextual analysis

First, there came to be much more detailed analysis of the contexts in which calls occurred than in previous investigations. For example, in his study of Japanese macaques, Green (1975) found that one call type, the coo call, actually consisted of several variants, each of which was associated with a different behavioral situation. In classical studies of primate vocalizations (e.g., Rowell and Hinde, 1962) data comprised a few representative sound spectrograms on the graded nature of calls. Actual isolation of discrete vocalizations based on physical characteristics was difficult because of this variability and because this variability was interpreted as representative of a behavioral continuum of arousal or motivation. In his study, Green therefore isolated additional sources of variability in the vocalizations of Japanese macaques. He sorted spectrograms into categories of similar appearing acoustic patterns and found that these categories represented vocalizations uttered in similar social contexts. Social contexts were differentiated by various factors such as age, biological state (e.g., "estrous female") and dominance relationships. His success in grouping calls according to their acoustic characteristics, which could then be correlated with social context, provided further support for the argument that vocalization variability is a function of behavioral categories.

Subsequently, for a number of vocalizations in other primate species that had been classified as single types, other researchers have found that an apparently unitary call type can further be divided into several variants (e.g., pygmy marmoset trills, Snowdon and Pola, 1978; cotton-top tamarin chirps and long calls, Cleveland and Snowdon, 1982). My own findings with Goeldi's monkey alarm calls provide another example. Examining the correlation between