In memory of Harriet Deacon, my grandmother and first mentor, who taught me to recognize the miraculous in everyday things.
Very special thanks are due to my family, Cris, Anneka, and John, who supported me even as I stole precious time away from them to complete this work, and to my parents, whose encouragement I could always count on. Thanks are also due to many who have directly inspired, assisted, and endured this project. These include: Joseph Marcus, who read, edited, and commented extensively on earlier drafts and who has consistently held me to the high standards he expects from a mentor; my own mentors, who have knowingly or unknowingly contributed their insights and valuable criticisms, but who are far too numerous to list; my former student Alan Sokoloff, whose Ph.D. research on oral tract innervation underlies my thoughts on vocal evolution; David Rudner, Sandra Kleinman, Alan Aronie, and many other friends whose feedback has helped clear away some of the fog; my colleagues in the lab, who understood and picked up more of the load as my attentions were diverted from the lab bench; Robyn Swierk and Julie Criniere, who helped enter the endless corrections, and copyeditor Ann Adelman; Hoover the seal, who opened my ears to the mystery of speech; and my editor, Hilary Hinzmann, who was patient through my long spells of writer's block and overcommitment, and who helped nurse a rough collection of ideas and notes into a narrative. I can never adequately repay these many gifts.
... the paradox is the source of the thinker's passion, and the thinker without a paradox is like a lover without feeling: a paltry mediocrity.

— Søren Kierkegaard

An Evolutionary Anomaly

As our species designation—*sapiens*—suggests, the defining attribute of human beings is an unparalleled cognitive ability. We think differently from all other creatures on earth, and we can share those thoughts with one another in ways that no other species even approaches. In comparison, the rest of our biology is almost incidental. Hundreds of millions of years of evolution have produced hundreds of thousands of species with brains, and tens of thousands with complex behavioral, perceptual, and learning abilities. Only one of these has ever wondered about its place in the world, because only one evolved the ability to do so.

Though we share the same earth with millions of kinds of living creatures, we also live in a world that no other species has access to. We inhabit a world
full of abstractions, impossibilities, and paradoxes. We alone brood about what didn’t happen, and spend a large part of each day musing about the way things could have been if events had transpired differently. And we alone ponder what it will be like not to be. In what other species could individuals ever be troubled by the fact that they do not recall the way things were before they were born and will not know what will occur after they die? We tell stories about our real experiences and invent stories about imagined ones, and we even make use of these stories to organize our lives. In a real sense, we live our lives in this shared virtual world. And slowly, over the millennia, we have come to realize that no other species on earth seems able to follow us into this miraculous place.

We are all familiar with this facet of our lives, but how, you might ask, could I feel so confident that it is not part of the mental experience of other species—so sure that they do not share these kinds of thoughts and concerns—when they cannot be queried about them? That’s just it! My answer, which will be argued in detail in the following chapters, has everything to do with language and the absence of it in other species. The doorway into this virtual world was opened to us alone by the evolution of language, because language is not merely a mode of communication, it is also the outward expression of an unusual mode of thought—symbolic representation. Without symbolization the entire virtual world that I have described is out of reach: inconceivable. My extravagant claim to know what other species could not know rests on evidence that symbolic thought does not come innately built in, but develops by internalizing the symbolic process that underlies language. So species that have not acquired the ability to communicate symbolically cannot have acquired the ability to think this way either.

The way that language represents objects, events, and relationships provides a uniquely powerful economy of reference. It offers a means for generating an essentially infinite variety of novel representations, and an unprecedented inferential engine for predicting events, organizing memories, and planning behaviors. It entirely shapes our thinking and the ways we know the physical world. It is so pervasive and inseparable from human intelligence in general that it is difficult to distinguish what aspects of the human intellect have not been molded and streamlined by it. To explain this difference and describe the evolutionary circumstances that brought it about are the ultimate challenges in the study of human origins.

The question that ultimately motivates a perennial fascination with human origins is not who were our ancestors, or how they came to walk upright, or even how they discovered the use of stone tools. It is not really a question that has a paleontological answer. It is a question that might otherwise be asked of psychologists or neurologists or even philosophers. Where do human minds come from? The missing link that we hope to fill in by investigating human origins is not so much a gap in our family tree, but a gap that separates us from other species in general. Knowing how something originated often is the best clue to how it works. And we know that human consciousness had a beginning. Those features of our mental abilities that distinguish us from all other species arose within the handful of million years since we shared a common ancestor with the remaining African apes, and probably can mostly be traced to events that took place only within the last 2 million. It was a Rubicon that was crossed at a definite time and in a specific evolutionary context. If we could identify what was different on either side of this divide—differences in ecology, behavior, anatomy, especially neuroanatomy—perhaps we would find the critical change that catapulted us into this unprecedented world full of abstractions, stories, and impossibilities, that we call human.

It is not just the origins of our biological species that we seek to explain, but the origin of our novel form of mind. Biologically, we are just another ape. Mentally, we are a new phylum of organisms. In these two seemingly incommensurate facts lies a conundrum that must be resolved before we have an adequate explanation of what it means to be human.

Advances in the study of human evolution, the brain, and language processes have led many scientists confidently to claim to be closing in on the final clues to this mystery. How close are we? Many lines of evidence seem to be converging on an answer. With respect to our ancestry, the remaining gaps in the fossil evidence of our prehistory are being rapidly filled in. Within the last few decades a remarkably rich picture of the sizes and shapes of fossil hominid bodies and brains has emerged. It is probably fair to say that at least with respect to the critical changes that distinguish us in this way from other apes, there are few missing links yet to be found, just particulars to be filled in. That crucial phase in hominid evolution when our ancestors’ brains began to diverge in relative size from other apes’ brains is well bracketed by fossils that span the range. As for the inside story, the neurosciences are providing powerful new tools with which it has become possible to obtain detailed images from working human brains performing language tasks, or to investigate the processes that build our brains during development and distinguish the brains of different species, or even to model neural processes outside of brains. Finally, linguists’ analyses of the logical structure of languages, their diversity and recent ancestry, and the patterns that characterize their development in children have provided a wealth of information about just what needs to be explained, and compar-
ative studies of animals’ communications in the wild and their language-like capacities in the laboratory have helped to frame these questions with explicit examples.

Despite all these advances, some critical pieces of the puzzle still elude us. Even though neural science has pried ever deeper into the mysteries of brain function, we still lack a theory of global brain functions. We understand many of the cellular and molecular details, we have mapped a number of cognitive tasks to associated brain regions, and we even have constructed computer simulations of networks that operate in ways that are vaguely like parts of brains; but we still lack insight into the general logic that ties such details together. On the whole, most neuroscientists take the prudent perspective that only by continuing to unmask the details of simple neural processes in simple brains, and slowly, incrementally, putting these pieces together, will we ever be able to address such global theoretical questions as the neural basis for language. We must add to this many new problems arising out of the comparisons of animal communication to language. If anything, these problems have become more complex and more confusing the more we have learned about the sophistication of other species’ abilities and the paradoxes implicit in our own abilities. But the most critical missing piece of the puzzle is access to the brains in question: ancestral hominid brains. Though we have considerable information about brain sizes in fossil species, and a little information about brain shapes, the relevant anatomical information, the internal microarchitecture of these brains, has left no fossil trail. With respect to fossil brains, we will never find the “smoking gun”—the first brain capable of language. We will only have access to circumstantial information.

So, what business do we have speculating about the beginnings of language? Given the complexity of the human brain, our current ignorance of many of its crucial operating principles, and the fact that neither languages nor the brains that produce them fossilize, there would appear to be many more immediate questions to be answered before even considering this one. There seem to be too many loose ends and gaps in the supportive evidence to provide solid leads in the search for clues to the nature of the human mind in the origins of language.

But this ignores the significance of the fact that language is a one-of-a-kind anomaly. Often the most salient and useful hints about the underlying logic of nature’s designs are provided when unique or extreme features in two different domains are found to be correlated. Some notable examples include the correlation between superconductivity and extreme cold; between greater cosmic distances and the increasing redness of starlight; between the massive extinctions of fossil species and evidence of extraterrestrial impacts; between the peculiarity of haplo-diploid genetics and war, suicidal defense, and infertile castes in social insects; and so on. Each of these correlations cried out for an explanation and in so doing offered a critical clue to a more general principle. The more two related features diverge from what is typical in their respective domains, the more penetrating the insight that can be gleaned from their underlying relationship.

In this context, then, consider the case of human language. It is one of the most distinctive behavioral adaptations on the planet. Languages evolved in only one species, in only one way, without precedent, except in the most general sense. And the differences between languages and all other natural modes of communicating are vast. Such a major shift in behavioral adaptation could hardly fail to have left its impression on human anatomy. Even superficial appearances bear this out. We humans have an anomalously large brain and a uniquely modified vocal tract. Though these clues offer no more than a starting point, they suggest that the structural and functional relationships underlying these superficial correlations are likely to be robust and idiosyncratic to us.

Ironically, then, the problem of language origins may actually offer one of the most promising entry points in the search for the logic linking cognitive functions to brain organization. To the extent that the unique mental demands of language are reflected in unique neuroanatomical differences, we may find an unequivocal example of how nature maps cognitive differences to brain structure differences. Though the details of this mystery are challenging, no critical pieces of this puzzle lie buried in the deep evolutionary past or inaccessible to current technology. They are observable in the differences in cognitive abilities and brain structures to be found in living species.

I think that the difficulty of the language origins question is not to be blamed on what we don’t know, but rather on what we think we already know. We think we know that what keeps language from being a widespread phenomenon is its byzantine complexity and the incredible demands it places on learning and memory. We think we know that language became possible for our ancestors when these impediments to language learning were overcome by some prior change in the brain. Depending on which aspects of language are deemed to be most complex, different prior adaptations are invoked to explain how language became possible. Perhaps it required an increase in intelligence, a streamlining of oral and auditory abilities, a separation of functions to the two sides of the brain, or the evolution of a sort of built-in grammar. I think we can be sure of none of these
things. In fact, I think that the problem is more basic and far more counterintuitive than is assumed by any of these accounts.

There are a few common assumptions shared by all of these explanations that I think are the root of a deeper problem. In general, these arguments parallel many others that continually resurface along that age old divide between nature and nurture. Is language imposed from the outside or does it reflect what is already inside? For decades, the superficiality of this stale dichotomy has been evident, exposed by research in the psychological and biological sciences that demonstrates how truly complex and interdependent the biological and environmental contributions to development can be; but we still find it difficult to conceive of these phenomena in other terms. We have reinvented the same old answers in new guises in each generation, stubbornly insisting that the answer to the question of language knowledge must be found in one of just a few major alternative paradigms (depicted in cartoon fashion in Figure 1.1).

At one end of this spectrum is the assumption that the architecture of language originates entirely outside (simple associationism); at the other end is the assumption that it originates entirely inside (mentalese). What other alternatives could there be, that are not captured between these extremes? And if there are no other alternatives, then shouldn't answering this question also point to the solution to the language origins question? Discovering which aspects of language knowledge are contributed by nature and which by nurture ought to tell us what difference in us was necessary to bridge the original language acquisition gap. If the answer lies more toward the associationist end of the spectrum, then evolution must have given us language by endowing us with exceptionally powerful learning and memory. If the answer lies more toward the mentalese end, then evolution must have endowed us with remarkably sophisticated instinctual knowledge of language that made learning completely unnecessary.

In light of these intuitively compelling alternatives, the approach I am about to take may seem misguided. Not only do I think that these alternatives confuse the nature/nurture problem more than they illuminate it, I think that the whole question of where language knowledge originates during development is secondary. Though a young child's almost miraculous development of language abilities is indeed a remarkable mystery—one that will be considered in some detail later (in Chapters 4 and 11)—I think that the cause of language origins must be sought elsewhere, and by pursuing some very different kinds of research questions. While we have been worrying about where knowledge of language comes from, we have been avoiding a more basic question: What sort of thing is knowledge of language

![Figure 1.1 Four cartoon depictions of some of the major theoretical paradigms proposed to explain the basis of human language. Top left: The notion that word meaning is created when the perception of the sound of a spoken word is associated with an object both as perceived and as stored in the mind in the form of a mental image. In this simple common sense view, stringing together words in a sentence leads the listener to bring together images in the mind. Top right: The notion that both word meaning and knowledge of language structure are learned by internalizing patterns of the associative probabilities linking words to one another, and linking words and objects. B. F. Skinner was the most prominent defender of this view, but recently more sophisticated versions of this basic idea have been reformulated with the aid of insights gained by studying parallel distributed learning processes. Knowledge of language is depicted as analogous to the distributed connection patterns in a neural net. Bottom left: One of the most influential views of grammatical knowledge conceives of it as built in prior to language experience, like firmware in a desktop computer (depicted as a computer chip inserted in the brain). The structure of language is imposed on strings of words (that presumably would still be meaningful, just less useful, without this structure). This view was first explicitly formulated by the linguist Noam Chomsky. Bottom right: The extreme innatist view of knowledge of language conceives of it as an external reflection of an internal lingua franca of the brain called "mentalese." In Steven Pinker's words (The Language Instinct, p. 82), "Knowing a language, then, is knowing how to translate mentalese into strings of words and vice versa. People without language would still have mentalese, and babies and many nonhuman animals presumably have simpler dialects. Indeed, if babies did not have a mentalese to translate to and from English, it is not clear how learning English could take place, or even what learning English would mean." None of these views provides a satisfactory explanation of the paradox explored in this chapter.
anyway? Before turning to this question, however, it is worth while reflecting on some of the equally misleading evolutionary assumptions that reinforce the traditional theoretical alternatives.

**Technical Difficulties and Hopeful Monsters**

One of the most common views about language evolution is that it is the inevitable outcome of evolution. Evolution was headed this way, our way. As the only species capable of conceiving of our place among all others, we see what looks like a continuous series of stages leading up to one species capable of such reflections. It goes without saying that a more articulate, more precise, more flexible means of communicating should always be an advantage, all other things being equal. In terms of cooperative behaviors, a better ability to pass on information about distant or hidden food resources, or to organize labor for a hunt, or to warn of impending danger, would be advantageous for kin and the social group as a whole. Better communication skills might also contribute to more successful social manipulation and deception. The ability to convince and mislead one's competitors or cooperate and connive with one's social and sexual partners could also have provided significant reproductive advantages, particularly in social systems where competition determines access to defendable resources or multiple mates. In fact, it's difficult to imagine any human endeavor that would not benefit from better communication. Looked at this way, it appears that humans have just developed further than other species along an inevitable progressive trend toward better thinking and better communicating.

Surely we must be part of a trend of better communication in some form? It seems to be an unstated assumption that if biological evolution continues long enough, some form of language will eventually evolve in many other species. Are chimpanzees the runners-up, lagging only a little behind on the road to language? As in Planet of the Apes, a science fiction movie in which our more hairy cousins catch up to a human level of mental and linguistic abilities, we imagine that if given sufficient time, something like language is prefigured in evolution. We even imagine that if there is life on other planets, and if it has been evolving as long as life on earth, or longer, there will be “intelligent” species with whom we may someday converse. The Renaissance notion of a “Great Chain of Being” gave rise to nineteenth-century theories of phylogeny that ranked species from lower to higher, from mechanism to godly, with humans just below angels. Though later nineteenth- and twentieth-century evolutionists rejected the static ranking of phylogeny and replaced it with the theory of evolutionary descent, the anthropocentric perspective was simply rephrased in evolutionary terms. Humans were presumably the most “advanced” species. Carrying this notion to its extreme, some people now suspect that there may be spaceships visiting earth, carrying beings that are “more highly evolved” than we are.

On the surface, progress seems to be implicit in natural selection. Gradual improvement of adaptations seems to imply that the longer evolution continues the better the design will become. Indeed, many scientists talk as though a special kind of retrograde selection would be necessary to halt the progress of inevitably increasing intelligence. Small-brained species are often considered primitive or throwbacks to earlier forms, left out of the main trend. From an anthropocentric perspective, it seems unquestionable that more intelligent species will outcompete less intelligent ones. Intelligence is always an advantage, right? Brain over brawn. We rank genius and mental retardation on a single scale, and presumably rank chimpanzees, dogs, and rats on the low end of the same scale. Human evolution is often termed an “ascent” to imply a climb from lower to higher intelligence. And from this it seems to follow that humans are just the pinnacle example of an inevitable trend. The winner in a war of neurons.

The apparent reasonableness of this view reflects our familiarity with technological progress in Western societies. The interchangeability of terms like consciousness expansion, social progress, and evolution is now almost commonplace in the popular press, and these ideas are seldom entirely disentangled even in the most sophisticated accounts of human evolution. But the idea of progress in evolution is an unnoticed habit left over from a misinformed common sense, from seeing the world in terms of design. The problem is that our intuitive model for evolution is borrowed from the history of technological change, which has been a cumulative process, adding more and more tidbits of know-how to the growing mass of devices, practices, and records each day. In contrast, biological evolution is not additive, except in some very limited ways. The human repertoire of expressed genes is about the same as that in a mouse or frog, and the body plans of all vertebrates seem to be mostly modifications of the same shared plan—even for the brain. Though we are on the large end of the range of body and brain sizes, this is not the result of adding new organs but merely enlarging existing ones with slight modifications.

Evolution is an irreversible process, a process of increasing diversification and distribution. Only in this sense does evolution exhibit a consistent direction. Like entropy, it is a process of spreading out to whatever possibilities are unfilled and within reach of a little more variation. Evolution does
not continue to churn out ever better mousetraps, even if it has produced some remarkable examples along the way. But this pattern of spreading into unfilled niches does place us in one of the more extreme niches.

Evolution is diversification in all directions, but there are more options available in some directions than others. Organisms started out small and short-lived and couldn’t get much smaller, but they could always get larger and more long-lived. For the smallest organisms, the resources that can be devoted to internal representations of the world are limited, though even bacteria appear able to use their one information-storage system, their genes, to take in information from around them and modify their behaviors appropriately. But the upper end of the range of information-handling abilities was not similarly bounded, and so the difference between the low end and the upper end of this range has increased over the hundreds of million years of animal evolution as part of this diversification. Nevertheless, the number of small-brained creatures has not diminished because of competition with those with big brains, and the no-brainers—all the plants and single-celled organisms—vastly outnumber the rest of us. It just happens that one very, very minor evolutionary direction is toward niches where doing a lot of information processing during one’s life is a good way to pass on genes. Inevitable? Well, it’s about as inevitable a direction in evolution as the development of arctic fish with antifreeze in their blood or electric eels who use electricity to sense their way through muddy Amazonian waters. The niche was just there, and was eventually filled. Still, in some measure, we are near the extreme of this distribution.

The question, then, is whether the evolution of language was somehow prefigured in this trend. Is there a general trend toward better communication? It’s easy in hindsight to arrange the history of long-distance communication from telegraphs to telephones to cellular phones to Star Trek communicators. It is not so easy to determine if animal communication has been steadily getting better and if human language is a part of such a trend. Certainly, there were advances in distance and signal clarity in evolution, but even if we narrow our comparison to vocal communication, there is no evidence in living species that some inevitable progressive trend leads to us. Apparently simple species can use highly complex methods of sound communication, and some highly complex species can be oblivious to their advantages. There are also many great sound tricks, such as echolocation, that are completely beyond human ability. Among our closest relatives, the great apes, there are both highly vocal (chimp) and nearly silent (orangutan) species. In fact, most birds easily outshine any mammal in vocal skills, and though dogs, cats, horses, and monkeys are remarkably capable learners in many domains, vocalization is not one of them. Our remarkable vocal abilities are not part of a trend, but an exception.

We also tend to underestimate the complexity and subtlety of much nonhuman social communication. In recent decades, field studies of social communication in nonhuman species have demonstrated that many birds, primates, and social carnivores use extensive vocal and gestural repertoires to structure their large social groups. These provide a medium for organizing foraging and group movement, for identifying individuals, for maintaining and restructuring multidimensional social hierarchies, for mediating aggressive encounters, for soliciting aid, and for warning of dangers. Indeed, even in our own species, a rich and complex language is still no substitute for a shocked expression, a muffled laugh, or a flood of silent tears, when it comes to communicating some of the more important messages of human social relationships.

However, although they are complex, these elaborate repertoires of calls, displays, and gestures do not seem to map onto any of the elements that compose languages. Although various researchers have suggested that parallels to certain facets of language are to be found in the learned dialects of birdsong, the external reference evident in vervet monkey alarm calls or honeybee dances, and the socially transmitted sequences of sounds that make up humpback whale songs (each of which will be considered in some detail in Chapter 2), these and many other examples like them only exhibit a superficial resemblance to language learning, word reference, or syntax, respectively. Even if we were to grant these parallels, no nonhuman species appears to put these facets of language together into a coordinated, rule-governed system.

Could we have missed recognizing nonhuman languages because they are as alien to us as our speech is to them? People have long entertained this possibility at least in mythology and children’s stories. They offer the fantasy that we might someday overcome the communication boundaries that separate humans and other animals and share memories, beliefs, hopes, and fears with them. In the popular children’s book made into a movie, Dr. Doolittle enlists the aid of a “multilingual” parrot to translate between animal and human speech. But is even a very superficial “translation” possible? What do you tell a child who asks, “What is the kitty saying?” Do animals’ vocalizations and gestures explain, describe, ask, or command? Do they argue, disagree, bargain, gossip, persuade, or entertain one another with their thoughts? Are there any corresponding elements in animal communication that map onto the elements of human language? Unfortunately, animal calls and displays have nothing that corresponds to noun parts or verb
parts of sentences, no grammatical versus ungrammatical strings, no marking of singular or plural, no indications of tense, and not even any elements that easily map onto words, except in the most basic sense of the beginning and ending of a sound.

One quite reasonable caution against making strong claims about the absence of nonhuman languages is that our study of other species’ communication systems is still in its infancy. Isn’t it more prudent to remain agnostic about the presence or absence of nonhuman languages? It is always wise not to prejudge the evidence, especially with respect to a subject about which we have undeniable prejudices. And there are far more species about whose communicative behavior we know next to nothing than there are whose communication has been studied. Nevertheless, I think that we have sufficient information to make a reasonably confident claim even about species whose communicative behaviors have been studied. What makes this a fairly safe guess is not the sophistication of our behavioral analyses, but rather the striking characteristics that would be evident in a nonhuman language. Where the differences should be glaring, the sensitivity and sophistication of observations and tests can be minimal.

What would be the characteristics of a nonhuman language that would allow us instantly to recognize it as a language-like form of communication, even if it were quite alien with respect to all human languages? This is the sort of question that scientists scanning the heavens with radio telescopes might ask by those engaged in electronic surveillance interested in distinguishing the transmission of coded or encrypted signals from random noise. Interestingly, many of these characteristics are exhibited in the surface structure of the signals, and require no special insight into meaning or referential function, and no obvious correspondence with natural language grammars, to discern. Here are a few general features that ought to stand out. A language-like signal would exhibit a combinatorial form in which distinguishable elements are able to recur in different combinations. It would exhibit a creative productivity of diverse outputs and a rather limited amount of large-scale redundancy. And although there would be a high degree of variety in the possible combinations of elements, the majority of combinatorial possibilities would be systematically excluded. In terrestrial examples, where it would be possible to observe the correlations between signals and contextual events, there should be another rather striking and counterintuitive feature. The correlations between individual signals and the objects and events that form the context of their production should not exhibit a simple one-to-one mapping. The correlations between sign elements and their pragmatic context should differ radically yet systematically from occasion to occasion, depending on how the signals are arranged in combination with respect to one another. These, of course, are all general features associated with syntax, though not just language syntax. Human games, mathematics, and even cultural customs exhibit these features.

If a radio-telescope observer identified a signal emanating from distant space with these characteristics it would make world headlines, despite the fact that the meaning of the signal would remain completely undecodable. With far more to go on than this, in even superficially studied animal communications, we can be reasonably sure that for the vast majority of likely candidate species such a signal has not yet been observed. Instead, though highly complex and sophisticated, the communicative behaviors in other species tend to occur as isolated signals, in fixed sequences, or in relatively unorganized combinations better described by summation than by formal rules. And their correspondences with events and behavioral outcomes, in the cases where this can be investigated, inevitably turn out to be of a one-to-one correlational nature. Though an as yet undescribed example of an animal communication system that satisfies these criteria cannot be ruled out, it seems reasonable to conclude that the chances are poor that it would have gone unobserved in common animal species, any more than we would miss it in a cosmic radio signal.

My point is not that we humans are better or smarter than other species, or that language is impossible for them. It is simply that these differences are not a matter of incommensurate kinds of language, but rather that these nonhuman forms of communication are something quite different from language. For this reason I think that the comparison is misguided, and useful only at a very superficial level. This fact should not be too difficult to appreciate because we all have personal experience of just the sort of incommensurability I am talking about. There are numerous human counterparts to other animal’s nonlinguistic communicative behaviors. We too have a wide range of innately produced and universally understood facial expressions, vocalizations, and gestures. As in other species, they are an irreplaceable component of human social communication. Yet this is not analogous to being bilingual. This other repertoire of communicative behaviors is not a language of gestures instead of words. It is something else. And although these human calls and gestures comprise an entirely familiar system, we find the same difficulty translating them into word equivalents as we do with animal calls and gestures with which we are far less familiar. The problem is not their unfamiliarity but rather that it simply makes no
sense to ask what kind of word a laugh is, whether a sob is expressed in past or present tense, or if a sequence of facial gestures is correctly stated. The problem isn’t a difficulty mapping human to nonhuman languages, but rather a difficulty mapping languages to any other form of naturally evolved communication, human or otherwise.

Of no other natural form of communication is it legitimate to say that “language is a more complicated version of that.” It is just as misleading to call other species’ communication systems simple languages as it is to call them languages. In addition to asserting that a Procrustean mapping of one to the other is possible, the analogy ignores the sophistication and power of animals’ nonlinguistic communication, whose capabilities may also be without language parallels. Perhaps we are predisposed to see other species’ communications through the filter of language metaphors because language is too much a natural part of our everyday cognitive apparatus to let us easily gain an outside perspective on it. Yet our experience of its naturalness, its matter-of-factness, belies its alien nature in the grander scheme of things. It is an evolutionary anomaly, not merely an evolutionary extreme.

This lack of precedent makes language a problem for biologists. Evolutionary explanations are about biological continuity, so a lack of continuity limits the use of the comparative method in several important ways. We can’t ask, “What ecological variable correlates with increasing language use in a sample of species?” Nor can we investigate the “neurological correlates of increased language complexity.” There is no range of species to include in our analysis. As a result, efforts to analyze the evolutionary forces responsible for language have often relied on crude substitutes to make up for the lack of homology between language and nonhuman forms of communication. It is tempting to try to conceive of language as an extrapolated extreme of something that other species produce, such as calls, grunts, gestures, or social grooming. It is also tempting to turn to some other feature of human anatomy that can be more easily compared to other species as an index of language evolution. Humans can, for example, be ranked along with other species with respect to brain size, group size, social-sexual organization, foraging strategy, etc. But even if humans are at the extreme in many of these measures, the correlations among these attributes are not obvious, and their linkage to language is dubious, since trends in these attributes in other species occur irrespective of language.

Interpreting the discontinuity between linguistic and nonlinguistic communication as an essential distinction between humans and nonhumans, however, has led to an equally exaggerated and untenable interpretation of language origins: the claim that language is the product of a unique one-of-a-kind piece of neural circuitry that provides all the essential features that make language unique (e.g., grammar). But this does not just assume that there is a unique neurological feature that correlates with this unique behavior, it also assumes an essential biological discontinuity. In other words, that language is somehow separate from the rest of our biology and neurology. It is as though we are apes plus language—as though one handed a language computer to a chimpanzee.

This reminds me of a wonderful piece of modern mythology from a recent film entitled Short Circuit. A sophisticated robot is accidentally transformed from a mechanism that “just runs programs” into a conscious, self-aware being as a result of being struck by lightning. The power surge damaged its circuits in just the right way. The now conscious robot, of course, does not think of this as “damage.” From his perspective, the lightning bolt corrected a design limitation. As a cinematic device, the bolt of lightning accomplishes two important things. The catastrophic and unpredictable nature of lightning provides a vehicle for invoking drastic and unprecedented change, and its intrinsically chaotic—and, by tradition, miraculous—character obviates any possibility of describing exactly what alterations changed a computer mechanism into a human-type mind. For the sake of the story, we suspend critical analysis and allow this miraculous accident to stand in place of an otherwise inexplicable transformation. As an allegory of human mental evolution, it offers a paradigm example of what biologists call a “hopeful monster” theory: the evolutionary theorist’s counterpart to divine intervention, in which a freak mutation just happens to produce a radically different and serendipitously better-equipped organism.

The single most influential “hopeful monster” theory of human language evolution was offered by the linguist Noam Chomsky, and has since been echoed by numerous linguists, philosophers, anthropologists, and psychologists. Chomsky argued that the ability of children to acquire the grammar of their first language, and the ability of adults effortlessly to use this grammar, can only be explained if we assume that all grammars are variations of a single generic “Universal Grammar,” and that all human brains come with a built-in language organ that contains this language blueprint. This is offered as the only plausible answer to an apparently insurmountable learning problem. Grammars appear to have an unparalleled complexity and systematic logical structure, the individual grammatical “rules” aren’t explicitly evident in the information available to the child, and when they acquire their first language children are still poor at learning many other things. Despite these limitations children acquire language knowledge at a remarkable rate.
This leads to the apparently inescapable conclusion that language information must already be "in the brain" before the process begins in order for it to be successful. Children must already "know" what constitutes an allowable grammar, in order to be able to ignore the innumerable false hypotheses about grammar that their limited experience might otherwise suggest.

This device, a "language organ" unique to the human brain, could also account for the failure of other species to acquire language. The appeal of this scenario is that it eliminates many troublesome questions in one fell swoop: the discontinuity between human and nonhuman communication, the larger human brain (adding a new part enlarges it), the systemic interdependent nature of grammatical rules (they all derive from one neurological source), the presumed universal features of language structure (ditto), the intertranslatability of languages (ditto), and the ease with which language is initially acquired despite an insufficient input and a lack of grammatical error correction by adults.

Another appeal of the hopeful monster story is that it promises a definite and dramatic transition from one stage to another in the evolutionary sequence. It offers a single-step evolutionary account that is much easier to comprehend and organize in one's thinking than continuous changes involving multiple factors interacting and overlapping in time in complex ways. It tantalizes the imagination to hear that the story of human origins was written in the course of a single dramatic and decisive prehistoric event. In one step, some ancestor crossed the threshold into humanity. But such a crucial transition could hardly occur without leaving a trail of evidence attesting to its discontinuity. If modern language abilities appeared all of a sudden in human prehistory, then we ought to find numerous other correlates of a radical reorganization of human behavior and biology. Inspired by this possibility, researchers in many fields have combed through their own data for signs of sudden transitions that might be the result of such an incredible language mutation. Not surprisingly, many have been "discovered" in the record of human prehistory. They include: abrupt technological transitions (e.g., the first appearance of stone tools or of extensive cultural variations in tool design); possible punctuated speciation events (e.g., the origination of anatomically modern humans from a "mitochondrial Eve"); rapid population changes (e.g., the demise of the Neanderthals); and signs of major innovations in cultural artefacts (the first appearance of durable representative art, such as carvings and cave paintings in Europe). But because they offer evidence that is indirect, at best, and so sparse and fragmentary, paleontological finds can appear irregular for many other reasons, not least of which is our predisposition to organize the evidence in categorical terms.

An accidental language organ requires no adaptive explanation for the structure of language. If this hypothetical organ was plugged into the brain in a single accident of prehistory, rather than evolving bit by bit with respect to its functional consequences, then no functional explanations would be necessary. If it was just an accident, any utility would be entirely accidental as well, discovered after the fact. This too might account for the many little idiosyncrasies of language and its discontinuities when compared with other nonhuman forms of communication. But I think that this story is far too neat and tidy precisely because it suggests that so many questions do not need to be addressed. The accidental language organ theory politely begs us to ignore the messy details of language origins, abandon hope of finding precedents in the structure of ape brains or their cognitive abilities, and stop looking for any deep design logic to the structural and functional relationships of language grammars and syntax. This is a lot to ignore. What does this hypothesis provide instead?

One of the characters in Molière's play The Imaginary Invalid is asked by his physician-examiners to explain the means by which opium induces sleep. He replies that it induces sleep because it contains a "soporific factor." This answer is applauded by the doctors. The playwright is, of course, satirizing the false expertise of these apparently learned men by showing their knowledge to be no more than sophistry. The answer is a nonexplanation. It merely takes what is in need of explanation and gives it a name, as though it were some physical object. Like phlogiston, the substance once hypothesized by pre-atomic chemistry to be the essence that determined flammability, the "soporific factor" fails to reduce the phenomenon in need of explanation to any more basic causal mechanisms.

For many linguists, grammatical knowledge is what needs to be explained, and what is lacking is an adequate account of the source of children's grammatical and syntactic abilities in terms of antecedents in the child's experience of language. We are thus like the characters in Molière's play, who know what is produced but don't know how it is produced. Failing to discover a satisfactory explanation for how grammatical knowledge could be impressed upon children's minds from the outside, we naturally turn to the possibility that it does not come from the outside at all. But simply assuming that this knowledge is already present, and so doesn't need to pass from outside to inside, only restates this negative finding in positive terms. A grammar instinct or a universal grammar serve as place holders for whatever could not be learned. The nature of this presumed innate
knowledge of language is described only in terms of its consequences. Linguists have progressively redefined what supposedly cannot be learned in ever more formal and precise terms, and so we may have the feeling that these accounts are approaching closer and closer to an explanation. But although the description of what is missing has gotten more precise, ultimately it is only a more and more precise version of what is missing. These “explanations” of the nature of a language instinct are inevitably presented in the guise of elaborate definitions of grammatical principles or else as something akin to computer programs, and in this way they are only more formal restatements of the problem of the missing information. Saying that the human brain alone produces grammar because it alone possesses a grammar factor ultimately passes the explanatory buck out of the hands of linguists and into the hands of neurobiologists.

To be fair, the intent of language organ theories is not to address the question of initial language origins, but rather to explain the source of language competence in development. For this reason, it is not wedded to the hopeful monster assumption. Steven Pinker, a proponent of the Universal Grammar view of language abilities and an articulate champion of many of Chomsky’s original insights about the uniqueness of language, argues in a recent book (The Language Instinct) that innate grammatical knowledge is not at all incompatible with an adaptationist interpretation of its origins. He argues that a language instinct could have gradually evolved through the action of natural selection. On the one hand, this is a far more biologically plausible alternative to miraculous accidents and it challenges us to face some of the difficult problems ignored by theories relying on miraculous accidents to fill in the gaps. On the other hand, an adequate formal account of language competence does not provide an adequate account of how it arose through natural selection, and the search for some new structures in the human brain to fulfill this theoretical vacuum, like the search for phlogiston, has no obvious end point. Failure to locate it in such a complex hierarchy of mechanisms can always be dismissed with the injunction: look harder.

A full evolutionary account cannot stop with a formal description of what is missing or a scenario of how selection might have favored the evolution of innate grammatical knowledge. It must also provide a functional account of why its particular organization was favored, how incremental and partial versions were also functional, and how structures present in nonhuman brains were modified to provide this ability. The language instinct theory provides an end point, an assessment of what a language evolution theory ultimately needs to explain. It rephrases the problem by giving it a new name. But this offers little more than the miraculous accident theory provided: a formal redescription of what remains unexplained. Unfortunately, I think it also misses the forest for the trees, even in this endeavor. I don’t think that children’s grammatical abilities are the crucial mystery of language.

The Missing Simple Languages

The two dominant paradigms for framing the language origins question—the evolution of greater intelligence versus the evolution of a specialized language organ—have one thing in common: both are stated in terms of the problem of learning a very large and complex set of rules and signs. They assume that other species are poor language learners because language is just too complicated for them to learn, and too demanding for them to perform. It requires rapid and efficient learning, demands immense memory storage, takes advantage of almost supernatural rates of articulation and auditory analysis, and poses an analytic problem that is worthy of a linguistic Einstein. Both approaches assume that the difficulty for other species is the complexity of language, but they disagree on the source of this difficulty and on what is required to overcome it. Do human children merely need to be very much more intelligent than other species in order to learn language, or is language so complicated that it is impossible to learn without some built-in language information to “jump start” the process? Accepting one or the other assumption leads to opposite claims about the nature of the evolution of language and the human mind. If language is just difficult to learn, then the neural adaptation that supports it could be quite general in its effect on cognitive abilities. If language is, for all practical purposes, impossible to learn, then the neural adaptation necessary to support it would need to be quite specific. However one looks at these problems, it appears that overcoming the limitations imposed by the obvious complexity of language is a prerequisite to language evolution. I say it “appears” this way, because I think that something has been missed in both views of the problem, something fundamental. These alternatives, and many plausible intermediates, only address one of the main problems in need of explanation, and it is not the critical one.

A task that is physically too difficult to perform may exceed our strength, our endurance, our rate of performance, our precision of action, our capacity to do many things at the same time, etc. In cognitive terms, these correspond to our ability to focus attention, the persistence of our memories, our rate of learning, and our short-term memory span, etc. When we say that
a skill is difficult to learn, we mean that the desired movement sequence severely taxes our ability precisely to time, control, or coordinate the component movements. When we say that a perceptual task is difficult to learn, we mean that it requires utilizing cues too subtle or fleeting to detect, too irregular to discover their commonalities, or embedded in too many distracting cues to sort out. And when we say that a cognitive task is difficult to learn, we mean that there are too many associations to be held in working memory at one time or too many to be considered in too little time or simply too many to be remembered. Each demands too much, from too little, in insufficient time. Both the complexity of the task and the resources one has available will determine its relative difficulty.

Clearly, language is complicated in all these ways. Linguistic communication requires us to learn and perform some remarkably complicated skills, both in the production of speech and in the analysis of speech sounds. In addition, there is a great deal to learn: thousands of vocabulary items and an intricate system of grammatical rules and syntactic operations. And it’s not enough that language is complicated. According to many linguists, we aren’t even offered sufficient outside support to deal with it. We are forced to figure out the underlying implicit rules of grammar and syntax without good teaching and with vastly inadequate examples and counterexamples. This apparent lack of adequate instruction adds insult to injury, so to speak, by making atoo complicated task even harder. The degree to which the support for language learning undershoots this need is (exponentially) proportional to how complicated the task is to begin with, and so the complexity of language is doubly limiting.

How could anyone doubt that language complexity is the problem? Languages are indeed complicated things. They are probably orders of magnitude more complicated than the next-most-complicated communication system outside of the human sphere. And they are indeed almost impossibly difficult for other species to acquire. The question is whether this complexity is the source of the difficulty that essentially limits the use of language to our species alone. Although this would seem to be the obvious conclusion, it is not quite so obvious as it might first appear. The most crucial distinguishing features of language cannot be accounted for merely in terms of language complexity.

The challenge to the complexity argument for human language origins rests on a simple thought experiment. Imagine a greatly simplified language, not a child’s language that is a fragment of a more complicated adult language, but a language that is logically complete in itself, but with a very limited vocabulary and syntax, perhaps sufficient for only a very narrow range of activities. I do not mean “language” in a metaphoric sense, the way that all communication systems are sometimes glossed as languages. But I also do not restrict my meaning to speech, or to a system whose organizational principles are limited to the sorts of grammatical rules found in modern languages. I mean language in the following very generic sense: a mode of communication based upon symbolic reference (the way words refer to things) and involving combinatorial rules that comprise a system for representing synthetic logical relationships among these symbols. Under this definition, manual signing, mathematics, computer “languages,” musical compositions, religious ceremonies, systems of etiquette, and many rule-governed games might qualify as having the core attributes of language. More important, no more than a tiny “vocabulary” of meaningful units and only two or three types of combinatorial rules would be necessary to fulfill these criteria. A five- or ten-word vocabulary and a syntax as simple as toddlers’ two- and three-word combinations would suffice. Reducing the definition of language to such minimal conditions allows us to conceive of languagelike systems that are far simpler even than the communicative repertoires found to occur in the social interactions of many other species.

So this is the real mystery. Even under these loosened criteria, there are no simple languages used among other species, though there are many other equally or more complicated modes of communication. Why not? And the problem is even more counterintuitive when we consider the almost unsurmountable difficulties of teaching language to other species. This is surprising, because there are many clever species. Though researchers report that languagelike communication has been taught to nonhuman species, both the successes and failures that have come of this research are never attests to the rather limited scope of the resulting behaviors, as well as to deep disagreements about what exactly constitutes languagelike behavior. Both the successes and failures that have come of this research are nevertheless highly informative with regard to both what animals can and can’t do and how we conceive of language itself, but the few arguable successes must be seen against the background of domesticated animals and family pets that never seem to catch on, despite being raised in a context where they are bombarded with a constant barrage of commands, one-sided conversations, and “rhetorical” questions. For the vast majority of species in the vast majority of contexts, even simple language just doesn’t compute. This lack of simple languages in the wild and inability to learn simple languages under human tutelage don’t make sense! Many of these species engage in natural communicative behaviors that are far more complex than a simple
language, and they are capable of learning larger sets of associations than are necessary for constructing a simple language. So why is language such a problem? The difference cannot be simple versus complex.

The complexity of language is important. It demands an explanation, as does the ability of young children to make sense of it, seemingly without sufficient feedback or time at their disposal. These are remarkable aspects of the language mystery, but they are secondary to a more basic mystery that has a lot more to do with the human/nonhuman difference. Despite the intelligence of other species, and the fact that they engage in communicative behaviors that are as complex in other ways as a simple language might be, no other language systems exist. And it's not just a matter of their not being needed. For some reason even a simple language seems impossibly difficult for nonhumans. This poses a profound riddle. So why has it been ignored? Perhaps we have been too preoccupied by the details to recognize this simpler problem. Or maybe we have been too eager to cast the problem in terms of progress in communication, with humans in the lead. Whatever the reason, it's time we recognized that the questions we thought needed to be explained by a theory of language origins were secondary to a more fundamental mystery: Why aren't there any simple languages? And why are even simple languages so nearly impossible for other species to learn?

This changes everything. If complexity is not the problem, then theories that purport to explain language evolution in terms of overcoming complexity lose their justification. A small vocabulary should not require vast intelligence or memory capacity or articulatory skill to master. Lower intelligence of our primate and mammalian relatives cannot, then, be the reason they don't catch on. A simple grammar and syntax should also be a trivial matter to learn. No special built-in encoder-decoder for grammars should be necessary if the combinatorial analyses are simple and the possible alternatives are relatively few. Even minimal powers of inductive learning would suffice. The whole raison d'être of an innate Universal Grammar or language organ evaporates when it comes to simple languages. Finally, complex phonology, rapid articulation, and automated speech sound analysis are equally unnecessary. The learning problems addressed by all these theories do not explain the absence of nonhuman languages, they only explain why nonhuman languages should not be as complicated as human ones. They point to issues that are relevant to complex modern human languages, but they do not illuminate the phenomenon we originally thought they explained. They don't provide any clue to why language evolved in the human lineage and nowhere else. This is an apples-versus-oranges problem, not a complicated-versus-simple one. It's not just curious that other species haven't started on this evolutionary path; it defies common sense.

What is left that is difficult about learning language, if its complexity is not at issue? When we strip away the complexity, only one significant difference between language and nonlanguage communication remains: the common, everyday miracle of word meaning and reference.

Neither grammar, nor syntax, nor articulate sound production, nor a huge vocabulary have kept other species from evolving languages. Just the simple problem of figuring out how combinations of words refer to things. Why should this be so difficult? Why should the curiously different way that languages represent things have imposed such an almost impenetrable barrier in evolution? If we succeed in explaining this one paradoxical difficulty, we may catch a glimpse of the critical evolutionary threshold that only our own ancestors managed to surmount.

The first major task of this book, then, is to describe precisely the difference between this unique human mode of reference, which can be termed symbolic reference, and the forms of nonsymbolic reference that are found in all nonhuman communication (and in many other forms of human communication as well). The second task is to explain why it is so incredibly difficult for other species to comprehend this form of reference. And the third task is to provide an explanation for how we humans (and a few other animals in carefully structured language learning experiments) have managed to overcome this difficulty. Even though this aspect of the language origins mystery is only a part of the story of language evolution, and understanding this difference offers no immediate answers to why languages are as complex as they are today, or why they obey seemingly inexplicable design rules, or how it is possible for human children to make sense of these otherwise byzantine and atypical details, none of these other questions can be answered without taking symbolic reference as a given. But it is not a given. Grammatical rules and categories are symbolic rules and categories. Syntactic structure is just physical regularity when considered irrespective of the symbolic operations it is intended to encode. Theories of language and mind that fail to address this issue head on, or suggest that it needs no explanation, ultimately assume what they set out to explain. We must explain the curious difficulty of symbolic reference first.

In hindsight, the centrality of this problem was recognized all along, at least implicitly. It would not be an exaggeration to suggest that more philosophical ink has been spilt over attempts to explain the basis for symbolic reference than over any other problem. Yet despite its intuitive familiarity (or because of it), and notwithstanding the efforts of some of the greatest
minds of each century, it remains curiously unresolved. Linguists, too, have struggled with this problem in the form of semantic theories, with parallel difficulties. For this reason, we should not be surprised to find that it resurfaces as the central riddle in the problem of language origins. Linguists, psychologists, and biologists cannot be blamed for failing to solve this basic mystery of mind before turning their efforts to other aspects of the language problem. Grammar and syntax can be studied and compared from language to language and the correlations between language processes and brain functions can also be identified irrespective of solving the problem of symbolic reference. Even many facets of language learning can be studied without considering it in any depth. But theories that purport to explain the difficult aspects of language cannot ignore it, nor can neuroanatomists and biologists. Human brains are unusually large: three times larger than they should be for an ape with our size body. But large brain size is only the most superficial symptom of a substantial reorganization at deeper levels. Unpacking this complicated anatomical problem and mapping it onto the special computational demands posed by language is the purpose of the middle section of this book. Looking more closely, we will discover that a radical re-engineering of the whole brain has taken place, and on a scale that is unprecedented. Interpreting these differences as consequences of the functional demands imposed by eons of language processing may offer new insight into the relationship between differences in cognitive function and differences in large-scale brain organization. In the co-evolution of the brain and language two of the most formidable mysteries of science converge, and together they provide a substantial set of clues about their relationship to one another.

Though neuroanatomists have been searching for a "Rosetta Stone" of human brain function for centuries, it has been far from a trivial task to sort out the significant from the incidental differences in brain structure and it
will take considerable effort just to identify exactly what has changed and how. As with the problem of determining what is fundamentally different about language, an analysis of how the brain has responded to these influences will require us to delve well below brain sizes and superficial differences in brain structures, to probe the processes that build brains in embryos. Brains are the most intricate and powerful computing devices on the planet. Linguistic communication is the most complex behavior known. Evolution is the epitome of inscrutability, indirectness, and opportunism—seldom following an obvious or elegant path. Now we must throw into this already daunting mix of problems the equally perplexing problem of explaining symbolic reference. A puzzle of such magnitude is unlikely to have an easy solution, nor do I imagine that what little evidence we have is sufficient to do any more than begin the process of hunting in the right places for more clues. But arranging the clues in the appropriate order is a first step, and considering an old problem from a novel perspective is often the best way to escape the maze of assumptions that prevents us from recognizing the obvious. Perhaps by juxtaposing these linked mysteries from different domains we will recognize the common thread of logic that runs through them all. Like the famous Rosetta Stone, on which the same text was written in three radically different scripts, these pieces of the cognitive and neural puzzle, aligned side by side, may enable us to discover how each translates the other.

If the only tool you have is a hammer, you tend to treat everything as if it were a nail.
—Abraham Maslow

Gymnastics of the Mind

We often find that an apparently simple task is difficult, not because it is complicated but because it is awkward. Sometimes, for instance, you just lack the right tools for the job. No matter how easy it is in principle to tighten a screw, if it has a slotted head (-) and all you have is a Phillips screwdriver (+), forget it. This is a familiar source of difficulty in physical activities. We humans are unprepared to perform a number of tasks that other species perform with ease. Other species can boast such adaptations as streamlining and fins for swimming, large flexible mobile surfaces for gliding and flying, claws for clinging to the trunks of trees or snagging prey, or sharp canines for tearing flesh. Attempt a task that you are poorly
spondence when we compare it to each of the examples above. If my use of the word “skunk” to refer to a certain animal was sustained by this crit­
ter being present, even a small percentage of the times that I used the word (in other words if there had to be a physical correlation), then the associa­
tion would have been extinguished long ago. A learned association will tend to get weaker and weaker if some significant degree of co-occurrence of stimuli is not maintained. I very seldom find myself in the company of members of this species, if I can help it, and yet I read and talk about them often. Despite this, I don’t have the impression that the strength of the refer­
tential link between the animals and the name is any less strong than that between the word “finger” and my flesh-and-blood finger, which is always present. There is some kind of word-object correspondence, but it isn’t based on a physical correlational relationship.

To understand this difference, then, we need to be able to describe the difference between the interpretive responses that are capable of sustain­
ing associations between a word and its reference, irrespective of their being correlated in experience, and those rote associations that are estab­
lished and dissolved as experience dictates. When we interpret the meaning and reference of a word or sentence, we produce something more than what a parrot produces when it requests a cracker or what a dog produces when it interprets a command. This “something more” is what constitutes our symbolic competence.

Alice laughed. “There’s no use trying,” she said: “one can’t believe impossible things.”

“I daresay you haven’t had much practice,” said the Queen.

“When I was your age I always did it for half-an-hour a day. Why, sometimes I’ve believed as many as six impossible things before breakfast.”

—Lewis Carroll, Alice Through the Looking-Glass

The Hierarchical Nature of Reference

The assumption that a one-to-one mapping of words onto objects and vice versa is, the basis for meaning and reference was made explicit in the work of the turn-of-the-century French linguist Ferdinand de Saussure. In his widely influential work on semiology (his term for the study of language), he argued that word meaning can be modeled by an element­by-element mapping between two “planes” of objects: from elements constituting the plane of the signifiers (e.g., words) to elements on the plane of the signified (the ideas, objects, events, etc., that words refer to). On this view, the mapping of vervet monkey alarm calls onto predators could be considered a signifier-signified relationship. But how accurately does this model word reference? Although it is natural to imagine words as labels for ob-
jects, or mental images, or concepts, we can now see that such correspondences only capture superficial aspects of word meaning. Focusing on correspondence alone collapses a multileveled relationship into a simple mapping relationship. It fails to distinguish between the rote understanding of words that my dog possesses and the semantic understanding of them that a normal human speaker exhibits. We also saw that the correspondence of words to referents is not enough to explain word meaning because the actual frequency of correlations between items on the two planes is extremely low. Instead, what I hope to show is that the relationship is the reverse of what we commonly imagine. The correspondence between words and objects is a secondary relationship, subordinate to a web of associative relationships of a quite different sort, which even allows us reference to impossible things.

In order to be more specific about differences in referential form, philosophers and semioticians have often distinguished between different forms of referential relationships. Probably the most successful classification of representational relationships was, again, provided by the American philosopher Charles Sanders Peirce. As part of a larger scheme of semiotic relationships, he distinguished three categories of referential associations: icon, index, and symbol. These terms were, of course, around before Peirce, and have been used in different ways by others since. Peirce confined the use of these terms to describing the nature of the formal relationship between the characteristics of the sign token and those of the physical object represented. As a first approximation these are as follows: icons are mediated by a similarity between sign and object, indices are mediated by some physical or temporal connection between sign and object, and symbols are mediated by some formal or merely agreed-upon link irrespective of any physical characteristics of either sign or object. These three forms of reference reflect a classic philosophical trichotomy of possible modes of associative relationship: (a) similarity, (b) contiguity or correlation, and (c) law, causality, or convention. The great philosophers of mind, such as John Locke, David Hume, Immanuel Kant, Georg Wilhelm Friedrich Hegel, and many others, had each in one way or another argued that these three modes of relationship describe the fundamental forms by which ideas can come to be associated. Peirce took these insights and rephrased the problem of relationship by studying the sign production and interpretation processes in more overt communication.

To get a sense of this logic of signs, let's begin by considering a few examples. When we say something is "iconic" of something else we usually mean that there is a resemblance that we notice. Landscapes, portraits, and pictures of all kinds are iconic of what they depict. When we say something is an "index" we mean that it is somehow causally linked to something else, or associated with it in space or time. A thermometer indicates the temperature of water, a weathervane indicates the direction of the wind, and an disagreeable odor might indicate the presence of a skunk. Most forms of animal communication have this quality, from pheromonal odors (that indicate an animal's physiological state or proximity) to alarm calls (that indicate the presence of a dangerous predator). Finally, when we say something is a "symbol," we mean there is some social convention, tacit agreement, or explicit code which establishes the relationship that links one thing to another. A wedding ring symbolizes a marital agreement; the typographical letter "e" symbolizes a particular sound used in words (or sometimes, as in English, what should be done to other sounds); and taken together, the words of this sentence symbolize a particular idea or set of ideas.

No particular objects are intrinsically icons, indices, or symbols. They are interpreted to be so, depending on what is produced in response. In simple terms, the differences between iconic, indexical, and symbolic relationships derive from regarding things either with respect to their form, their correlations with other things, or their involvement in systems of conventional relationships.

When we apply these terms to particular things, for instance, calling a particular sculpture an icon, a speedometer an indicator, or a coat of arms a symbol, we are engaging in a sort of tacit shorthand. What we usually mean is that they were designed to be interpreted that way, or are highly likely to be interpreted that way. So, for example, a striking resemblance does not make one thing an icon of another. Only when considering the features of one brings the other to mind because of this resemblance is the relationship iconic. Similarity does not cause iconicity, nor is iconicity the physical relationship of similarity. It is a kind of inferential process that is based on recognizing a similarity. As critics of the concept of iconicity have often pointed out, almost anything could be considered an icon of anything else, depending on the vagueness of the similarity considered.

The same point can be made for each of the other two modes of referential relationship: neither physical connection nor involvement in some conventional activity dictates that something is indexical or symbolic, re-
spectively. Only when these are the basis by which one thing invokes another are we justified in calling their relationship indexical or symbolic. Though this might seem an obvious point, confusion about it has been a source of significant misunderstandings. For example, there was at one time considerable debate over whether hand signs in American Sign Language (ASL) are iconic or symbolic. Many signs seemed to resemble pantomime or appeared graphically to “depict” or point to what was represented, and so some researchers suggested that their meaning was “merely iconic” and by implication, not wordlike. It is now abundantly clear, however, that despite such resemblances, ASL is a language and its elements are both symbolic and wordlike in every regard. Being capable of iconic or indexical interpretation in no way diminishes these signs’ capacity of being interpreted symbolically as well. These modes of reference aren’t mutually exclusive alternatives; though at any one time only one of these modes may be prominent, the same signs can be icons, indices, and symbols depending on the interpretive process. But the relationships between icons, indices, and symbols are not merely a matter of alternative interpretations. They are to some extent internally related to one another.

This is evident when we consider examples where different interpreters are able to interpret the same signs to a greater or lesser extent. Consider, for example, an archeologist who discovers some elaborate markings on clay tablets. It is natural to assume that these inscriptions were used symbolically by the people who made them, perhaps as a kind of primitive writing. But the archeologist, who as yet has no Rosetta Stone with which to decode them, cannot interpret them symbolically. The archeologist simply infers that to someone in the past these may have been symbolically interpretable, because they resemble symbols seen in other contexts. Being unable to interpret them symbolically, he interprets them iconically. Some of the earliest inscription systems from the ancient Middle Eastern civilizations of the Fertile Crescent were in fact recovered in contexts that provided additional clues to their representations. Small clay objects were marked with repeated imprints, then sealed in vessels that accompanied trade goods sent from one place to another. Their physical association with these other artifacts has provided archeologists with indexical evidence to augment their interpretations. Different marks apparently indicated a corresponding number of items shipped, probably used by the recipient of the shipment to be sure that all items were delivered. No longer merely iconic of other generic writinglike marks, they now can be given indexical and tentative symbolic interpretations, because something more than resemblance is provided.

This can also be seen by an inverse example: a descent down a hierar-

chy of diminishing interpretive competence, but this time with respect to interpretive competences provided by evolution. Let’s consider laughter again. Laughter indicates something about what sort of event just preceded it. As a symptom of a person’s response to certain stimuli, it provides considerable information about both the laugher and the object of the laugher, i.e., that it involved something humorous. But laughter alone does not provide sufficient information to reconstruct exactly what was so funny. Chimpanzees also produce a call that is vaguely similar to laughter in certain play situations (e.g., tickling). Consequently, they might also recognize human laughter as indicating certain aspects of the social context (i.e., playful, nonthreatening, not distressing, etc.), but they would likely miss the reference to humor. I suspect that implicit in the notion of humor there is a symbolic element, a requirement for recognizing contradiction or paradox, that the average chimpanzee has not developed. The family cat and dog, however, probably do not even get this much information from a human laugh. Not sharing our evolved predisposition to laugh in certain social relationships, they do not possess the mental prerequisites to interpret even the social signaling function of laughter. Experience may only have provided them with the ability to use it as evidence that a human is present and is probably not threatening. Nevertheless, this too is dependent on some level of interpretative competence, perhaps provided by recalling prior occasions when some human made this odd noise. Finally, there are innumerable species of animals from flies to snails to fish that wouldn’t even produce this much of a response, and would interpret the laughter as just another vibration of the air or water. The diminishing competences of these species corresponds with interpretations that are progressively less and less specific and progressively more and more concrete. But even at the bottom of this descent there is a possibility of a kind of minimalistic reference.

This demonstrates one of Peirce’s most fundamental and original insights about the process of interpretation: the difference between different modes of reference can be understood in terms of levels of interpretation. Attending to this hierarchical aspect of reference is essential for understanding the difference between the way words and animal calls are related. It’s not just the case that we are able to interpret the same sign in different ways, but more important, these different interpretations can be arranged in a sort of ascending order that reflect a prior competence to identify higher-level associative relationships. In other words, reference itself is hierarchical in structure; more complex forms of reference are built up from simpler forms. But there is more to this than just increasing complexity. This hierarchical structure is a clue to the relationships between these different
modes of reference. Though I may fail to grasp the symbolic reference of a sign, I might still be able to interpret it as an index (i.e., as correlated with something else), and if I also fail to recognize any indexical correspondences, I may still be able to interpret it as an icon (i.e., recognize its resemblance to something else). Breakdown of referential competence leads to an ordered descent from symbolic to indexical to iconic, not just from complex icons, indices, or symbols to simpler counterparts. Conversely, increasing the sophistication of interpretive competence reverses the order of this breakdown of reference. For example, as human children become more competent and more experienced with written words, they gradually replace their iconic interpretations of these marks as just more writing with indexical interpretations supported by a recognition of certain regular correspondences to pictures and spoken sounds, and eventually use these as support for learning to interpret their symbolic meanings. In this way they trace a path somewhat like the archeologist learning to decipher an ancient script.

This suggests that indexical reference depends upon iconic reference, and symbolic reference depends upon indexical reference—a hierarchy diagrammatically depicted in Figure 3.1. It sounds pretty straightforward on the surface. But this simplicity is deceiving, because what we really mean is that the competence to interpret something symbolically depends upon already having the competence to interpret many other subordinate relationships indexically, and so forth. It is one kind of competence that grows out of and depends upon a very different kind of competence. What constitutes competence in this sense is the ability to produce an interpretive response that provides the necessary infrastructure of more basic iconic and/or indexical interpretations. To explain the basis of symbolic communication, then, we must describe what constitutes a symbolic interpretant, but to do this we need first to explain the production of iconic and indexical interpretants and then to explain how these are each recoded in turn to produce the higher-order forms.

So, we need to start the explanation of symbolic competence with an explanation of what is required in order to interpret icons and build upward. Usually, people explain icons in terms of some respect or other in which two things are alike. But the resemblance doesn’t produce the iconicity. Only after we recognize an iconic relationship can we say exactly what we saw in common, and sometimes not even then. The interpretive step that establishes an iconic relationship is essentially prior to this, and it is something negative, something that we don’t do. It is, so to speak, the act of not making a distinction. Let me illustrate this with a very stripped-down example.

Consider camouflage, as in the case of natural protective coloration. A moth on a tree whose wings resemble the graininess and color of the bark, though not perfectly, can still escape being eaten by a bird if the bird is inattentive and interprets the moth’s wings as just more tree. Admittedly, this is not the way we typically use the term iconic, but I think it illuminates the most basic sense of the concept. If the moth had been a little less matching, or had moved, or the bird had been a little more attentive, then any of the differences between the moth and the tree made evident by those additional differences would have indicated to the bird that there was something else present which wasn’t just more tree. If the bird had been in a contemplative mood, it might even have reflected on the slight resemblance of the wing pattern to bark, at least for the fraction of a second before it gobbled the hapless moth. Some features of the moth’s wings were iconic of the bark, irrespective of their degree of similarity, merely because under some interpretation (an inattentive bird) they were not distinguished from it.

Now, it might seem awkward to explain iconicity with an example that could be considered to be no representation at all, but I think it helps to clarify the shift in emphasis I want to make from the relationship to the
process behind it. What makes the moth wings iconic is an interpretive process produced by the bird, not something about the moth's wings. Their coloration was taken to be an icon because of something that the bird didn't do. What the bird was doing was actively scanning bark, its brain seeing just more of the same (bark, bark, bark . . . ). What it didn't do was alter this process (e.g., bark, bark, not-bark, bark . . . ). It applied the same interpretive perceptual process to the moth as it did to the bark. It didn't distinguish between them, and so confused them with one another. This established the iconic relationship between moth and bark. Iconic reference is the default. Even in an imagined moment of reflective reverie in which the bird ponders on their slight resemblance, it is the part of its responding that does not distinguish wing from bark that determines their relationship to be iconic. Iconic resemblance is not based on some prior ground of physical similarity, but in that aspect of the interpretation process that does not differ from some other interpretive process. Thus, although a respect in which two things are similar may influence the ways they tend to be iconically related, it does not determine their iconicity. Iconism is where the referential buck stops when nothing more is added. And at some level, due either to limitations in abilities to produce distinguishing responses or simply a lack of effort to produce them, the production of new interpretants stops. Whether because of boredom or limitations of a minimal nervous system, there are times when almost anything can be iconic of anything else (stuff, stuff, stuff . . . ).

What does this have to do with pictures, or other likenesses such as busts or caricatures that we more commonly think of as icons? The explanation is essentially no different. That facet or stage of my interpretive recognition process that is the same for a sketch and the face it portrays is what makes it an icon. I might abstractly reflect on what aspects of the sketch caused this response, and might realize that this was the intention of the artist, but a sketch that is never seen is just paper and charcoal. It could also be interpreted as something that soaked up spilled coffee (and the spilled coffee could be seen as a likeness of Abe Lincoln!). Peirce once characterized an icon as something which upon closer inspection can provide further information about the attributes of its object. Looking at the one is like looking at the other in some respects. Looking at a caricature can, for example, get one to notice for the first time that a well-known politician has a protruding jaw or floppy jowls. The simplification in a diagram or the exaggeration in a cartoon takes advantage of our spontaneous laxness in making distinctions to trick us into making new associations. In this way a caricature resembles a joke, a visual pun, and a diagram can be a source of discovery.

In summary, the interpretive process that generates iconic reference is none other than what in other terms we call recognition (mostly perceptual recognition, but not necessarily). Breaking down the term re-cognition says it all: to "think [about something] again." Similarly, representation is to present something again. Iconic relationships are the most basic means by which things can be re-presented. It is the base on which all other forms of representation are built. It is the bottom of the interpretive hierarchy. A sign is interpreted, and thus seen to be a representation, by being reduced (i.e., analyzed to its component representations) to the point of no further reducibility (due to competence or time limitations, or due to pragmatic constraints), and thus is ultimately translated into iconic relationships. This does not necessarily require any effort. It is in many cases where interpretive effort ceases. It can merely be the end of new interpretation, that boundary of consciousness where experience fades into redundancy.

Interpreting something as an indexical relationship is this and more. Physical contiguity (nearness or connectedness) or just predictable co-occurrence are the basis for interpreting one thing as an index for another, but as with the case of icons, these physical characteristics are not the cause of the indexical relationship. Almost anything could be physically or temporally associated with anything else by virtue of some extension of the experience of nearness in space or time. What makes one an index of another is the interpretive response whereby one seems to "point to" the other. To understand the relationship that indexical interpretations have to iconic interpretations, it is necessary to see how the competence to make indexical interpretations arises. In contrast to iconic interpretations, which can often be attributed to interpretive incompetence or the cessation of production of new interpretants, indexical interpretations require something added. In fact, icons arise from a failure to produce critical indices to distinguish things.

Consider the example of a symptom, like the smell of smoke. When I smell smoke, I begin to suspect that something is burning. How did my ability to treat this smell as an indication of fire arise? It likely arose by learning, because I had past experiences in which similar odors were traced to things that were burning. After a few recurrences it became a familiar association, and the smell of smoke began to indicate to me that a fire might be there. If we consider more closely the learning process that produced the indexical competence, the critical role of icons becomes obvious. The indexical competence is constructed from a set of relationships between icons, and the indexical interpretation is accomplished by bringing this assembly of iconic relationships to bear in the assessment of new stimuli. The
smell of smoke brings to mind past similar experiences (by iconically representing them). Each of these experiences comes to mind because of their similarities to one another and to the present event. But what is more, many of these past experiences also share other similarities. On many of these occasions I also noticed something burning that was the source of the smoke, and in this way those experiences were icons of each other.

There is one important feature added besides all these iconic recognitions. The repeated correlation between the smelling of smoke and the presence of flames in each case adds a third higher-order level of iconicity. This is the key ingredient. Because of this I recognize the more general similarity of the entire present situation to these past ones, not just the smoke and not just the fire but also their co-occurrence, and this is what brings to mind the missing element in the present case: the probability that something is burning. What I am suggesting, then, is that the responses we develop as a result of day-to-day associative learning are the basis for all indexical interpretations, and that this is the result of a special relationship that develops among iconic interpretive processes. It's hierarchic. Prior iconic relationships are necessary for indexical reference, but prior indexical relationships are not in the same way necessary for iconic reference. This hierarchic dependency of indices on icons is graphically depicted in Figure 3.2.

Okay, why have I gone to all this trouble to rename these otherwise common, well-established uses of perception and learning? Could we just substitute the word “perception” for “icon” and “learned” association for index? No. Icons and indices are not merely perception and learning, they refer to the inferential or predictive powers that are implicit in these neural processes. Representational relationships are not just these mechanisms, but a feature of their potential relationship to past, future, distant, or imaginary things. These other things are not physically re-presented but only virtually re-presented by producing perceptual and learned responses like those that would be produced if they were present. In this sense, mental processes are no less representational than external communicative processes, and communicative processes are no less mental in this regard. Mental representation reduces to internal communication.

What, then, is the difference between these uncontroversial cognitive processes underlying icons and indices and the kind of cognitive processes underlying symbols? The same hierarchical logic applies. As indices are constituted by relationships among icons, symbols are constituted by relationships among indices (and therefore also icons). However, what makes this a difficult step is that the added relationship is not mere correlation.

The Symbolic Threshold

The common sense idea is that a symbolic association is formed when we learn to pair a sound or typed string with something else in the world. But in the terms we have been developing, this is what we mean by an indexical association. The word (iconically associated with past occurrences of similar utterances) and the object (iconically associated with similar ob-

Figure 3.2 A schematic diagram depicting the internal hierarchic relationships between iconic and indexical reference processes. The probability of interpreting something as iconic of something else is depicted by a series of concentric domains of decreasing similarity and decreasing iconic potential among objects. Surrounding objects have a decreasing capacity to serve as icons for the target object as similarities become unobvious. The form of a sign stimulus (S) elicits awareness of a set of past stimulus memories (e.g., mental “images”) by virtue of stimulus generalization processes. Thus, any remembered object (O) can be said to be re-presented by the iconic stimulus. Similarly, each mental image is iconic in the same way; no other referential relationship need necessarily be involved for an iconic referential relationship to be produced. Indexical reference, however, requires iconic reference. In order to interpret something as indexical, at least three iconic relationships must be also recognized. First, the indicating stimulus must be seen as an icon of other similar instances (the top iconic relationships); second, instances of its occurrence must also correlate (arrows) with additional stimuli either in space or time, and these need to be iconic of one another (the bottom iconic relationships); and third, past correlations need to be interpreted as iconic of one another (indicated by the concentric arrangement of arrows). The indexical interpretation is thus the conjunction of three iconic interpretations, with one being a higher-order icon than the other two (i.e., treating them as parts of a whole). As pointed out in the text, this is essentially the kind of reference provided by a conditioned response.
objects from past experiences) and their past correlations enable the word to bring the object to mind. In this view, the association between a word and what it represents is not essentially distinguished from the kind of association that is made by an animal in a Skinner box. We might, for example, train a rat to recognize a correlation between hearing the sound of the word "food" and food being dropped into a tray. The conditioned stimulus takes on referential power in this process: it represents something about the state of the apparatus for the animal. It is an index of the availability of food in the Skinner box; a symptom of the state of the box. Words can serve indexical functions as well, and are sometimes used for this purpose almost exclusively, with minimal symbolic content. Consider, for example, the use of function words like "there," exclamations like "Aha!" or even proper names like "George Washington." These derive reference by being uniquely linked to individual contexts, objects, occasions, people, places, and so on, and they defy our efforts to define them as we would typical nouns or verbs.

One indication that someone understands the meaning of a new word is whether they can use it in a new sentence or novel context. If the new word was just learned as a part of an unanalyzed phrase, or mapped to some restricted acquisition context, then we might not expect it to be correctly used out of this context. But the ability to use a word correctly in a variety of contexts, while fair evidence of symbolic understanding, is not necessarily convincing as a proof of understanding. The ability to shift usage to a novel context resembles transference of one learning set; and indeed, searching for the common learning set features among the many contexts in which the same word might be used is a good way to zero in on its meaning. If someone were to learn only this—i.e., that a particular phrase works well in a range of contexts that exhibit similar features or social relationships—they might well be able to fool us into believing that they understood what they said. However, on discovering that they accomplished this by simply mapping similar elements from one context to another, we would conclude that they actually did not understand the word or its role in context in the way we originally imagined. Theirs would be an iconic and indexical understanding only. Being able easily to transfer referential functions from one “set” to another is a characteristic of symbols, but is this the basis for their reference?

Psychologists call transfer of associations from one stimulus to another similar one “stimulus generalization,” and transfer of a pattern of learning from one context to another similar context the transfer of a “learning set.” These more complex forms of indexical association are also often confused with symbolic associations. Transference of learning from stimulus to stimulus or from context to context occurs as an incidental consequence of learning. These are not really separate forms of learning. Both are based on iconic projection of one stimulus condition onto another. Each arises spontaneously because there is always some ambiguity as to what are the essential parameters of the stimulus that a subject learns to associate with a subsequent desired or undesired result: learning is always an extrapolation from a finite number of examples to future examples, and these seldom provide a basis for choosing between all possible variations of a stimulus. To the extent that new stimuli exhibit features shared by the familiar set of stimuli used for training, and none that are inconsistent with them, these other potential stimuli are also incidentally learned. Often, psychological models of this process are presented as though the subject has learned rules for identifying associative relationships. However, since this is based on an iconic relationship, there is no implicit list of criteria that is learned; only a failure to distinguish that which hasn’t been explicitly excluded by the training.

Words for kinds of things appear to refer to whole groups of loosely similar objects, such as could be linked by stimulus generalization, and words for qualities and properties of objects refer to the sorts of features that are often the basis for stimulus generalization. Animals can be trained to produce the same sign when presented with different kinds of foods, or trees, or familiar animals, or any other class of objects that share physical attributes in common, even subtle ones (e.g., all hoofed mammals). Similarly, the vervet monkeys’ eagle alarm calls might become generalized to other aerial predators if they were introduced into their environment. The grouping of these referents is not by symbolic criteria (though from outside we might apply our own symbolic criteria), but by iconic overlap that serves as the basis for their common indexical reference. Stimulus generalization may contribute essential structure to the realms to which words refer, but it is only one subordinate component of the relationship and not what determines their reference.

This same logic applies to the transference of learning sets. For example, learning to choose the odd-shaped object out of three, where two are more similar to each other than the third, might aid in learning a subsequent oddity-discrimination task involving sounds. Rather than just transferring an associated response on the basis of stimulus similarities, the subject recognizes an iconicity between the two learning tasks as wholes. Though this is a hierarchically more sophisticated association than stimulus generalization—learning a learning pattern—it is still an indexical association transferred to a novel stimulus via an iconic interpretation. Here the structure
of the new training context is seen as iconic of a previous one, allowing the subject to map corresponding elements from the one to the other. This is not often an easy association to make, and most species (including humans) will fail to discover the underlying iconicity when the environment, the training stimuli, the specific responses required, and the reinforcers are all quite different from one context to the next.

There are two things that are critically different about the relationships between a word and its reference when compared to transference of word use to new contexts. First, for an indexical relationship to hold, there must be a correlation in time and place of the word and its object. If the correlation breaks down (for example, the rat no longer gets food by pushing a lever when the sound “food” is played), then the association is eventually forgotten (“extinguished”), and the indexical power of that word to refer is lost. This is true for indices in general. If a smokelike smell becomes common in the absence of anything burning, it will begin to lose its indicative power in that context. For the Boy Who Cried Wolf, in the fable of the same name, the indexical function of his use of the word “wolf” fails because of its lack of association with real wolves, even though the symbolic reference remains. Thus, symbolic reference remains stable nearly independent of any such correlations. In fact, the physical association between a word and an appropriate object of reference can be quite rare, or even an impossibility, as with angels, unicorns, and quarks. With so little correlation, an indexical association would not survive.

Second, even if an animal subject is trained to associate a number of words with different foods or states of the box, each of these associations will have little effect upon the others. They are essentially independent. If one of these associations is extinguished or is paired with something new, it will likely make little difference to the other associations, unless there is some slight transference via stimulus generalization. But this is not the case with words. Words also represent other words. In fact, they are incorporated into quite specific individual relationships to all other words of a language. Think of the way a dictionary or thesaurus works. They each map one word onto other words. If this shared mapping breaks down between users (as sometimes happens when words are radically reused in slang, such as “bad” for “very good” or “plastered” for “intoxicated”), the reference also will fail.

This second difference is what ultimately explains the first. We do not lose the indexical associations of words, despite a lack of correlation with physical referents, because the possibility of this link is maintained implicitly in the stable associations between words. It is by virtue of this sort of dual reference, to objects and to other words (or at least to other semantic alternatives), that a word conveys the information necessary to pick out objects of reference. This duality of reference is captured in the classic distinction between sense and reference. Words point to objects (reference) and words point to other words (sense), but we use the sense to pick out the reference, not vice versa.

This referential relationship between the words—words systematically indicating other words—forms a system of higher-order relationships that allows words to be about indexical relationships, and not just indices in themselves. But this is also why words need to be in context with other words, in phrases and sentences, in order to have any determinate reference. Their indexical power is distributed, so to speak, in the relationships between words. Symbolic reference derives from combinatorial possibilities and impossibilities, and we therefore depend on combinations both to discover it (during learning) and to make use of it (during communication). Thus the imagined version of a nonhuman animal language that is made up of isolated words, but lacking regularities that govern possible combinations, is ultimately a contradiction in terms.

Even without struggling with the philosophical subtleties of this relationship, we can immediately see the significance for learning. The learning problem associated with symbolic reference is a consequence of the fact that what determines the pairing between a symbol (like a word) and some object or event is not their probability of co-occurrence, but rather some complex function of the relationship that the symbol has to other symbols. This is a separate but linked learning problem, and worse yet, it creates a third, higher-order unlearning problem. Learning is, at its base, a function of the probability of correlations between things, from the synaptic level to the behavioral level. Past correlations tend to be predictive of future correlations. This, as we’ve seen, is the basis for indexical reference. In order to comprehend a symbolic relationship, however, such indexical associations must be subordinated to relationships between different symbols. This is a troublesome shift of emphasis. To learn symbols we begin by learning symbol-object correlations, but once learned, these associations must be treated as no more than clues for determining the more crucial relationships. And these relationships are not highly correlated; in fact, often just the reverse. Words that carry similar referential function are more often used alternatively and not together, and words with very different (complementary) referential functions tend to to be adjacent to one another in sentences. Worst of all, few sentences or phrases are ever repeated exactly, and the frequency with which specific word combinations are repeated is also extremely low. Hardly a recipe for easy indexical learning.
One of the most insightful demonstrations of the learning difficulties associated with the shift from conditioned associations to symbolic associations comes not from a human example, but from a set of experiments that attempted to train chimpanzees to use simple symbols. This study was directed by Sue Savage-Rumbaugh and Duane Rumbaugh, now at the Language Research Center of Georgia State University, and included four chimps, two of which, Sherman and Austin, showed particular facility with the symbols. It is far from the "last word" on how far other species can go in their understanding of languagelike communication, and further studies of another chimpanzee (from a different subspecies) that show more developed abilities will be described subsequently (see Chapter 4), but this work has the virtue of exposing much of what is often hidden in children's comparatively easy entry into symbolic communication, and so provides an accessible step-by-step account of what we usually take for granted in the process. In what follows I will outline these experiments briefly. Only the most relevant highlights will be described and other aspects will be simplified for the sake of my purpose here. Of course, my attempts to "get inside the chimps' heads" during this process are fantasy. Though I will use somewhat different terminology from the experimenters to describe this transition from indexical to symbolic communication, I am reasonably confident that my interpretation is not at odds with theirs. However, the interested reader should refer to the excellent account of these experiments and their significance in Savage-Rumbaugh's book describing them.

The chimps in this study were taught to use a special computer keyboard made up of lexigrams—simple abstract shapes (lacking any apparent iconism to their intended referents) on large illuminated keys on a keyboard mounted in their cage. Duane Rumbaugh's previous experiments (with a chimp named Lana) had shown that chimps have the ability to learn a large number of paired associations between lexigrams (and in fact other kinds of symbol tokens) and objects or activities. But in order to respond to critics and more fully test other features of this ability, Duane and Sue began a new series of experiments with a group of chimps to test both chim-chimp communication and chimps' ability to use lexigrams in combinations (e.g., syntactic relationships). Not surprisingly, the chimps exhibited some interesting difficulties when they were required to use lexigrams in combinations, but they eventually solved their learning problems and exhibited a use of the lexigrams that was clearly symbolic. In so doing they have provided us with a remarkably explicit record of the process that leads from index to symbol.

In order to test Sherman and Austin's symbolic understanding of the lex-
immediately punished or at least explicitly not rewarded (the correlation problem again). So the chimps were first trained to produce incorrect associations (e.g., mistaking keyboard position as the relevant variable) and then these errors were explicitly not rewarded, whereas the remaining appropriate responses were. By a complex hierarchic training design, involving thousands of trials, it was possible to teach them to exclude systematically all inappropriate associative and combinatorial possibilities among the small handful of lexigrams. At the end of this process, the animals were able to produce the correct lexigram strings every time.

Had training out the errors worked? To test this, the researchers introduced a few new food items and corresponding new lexigrams. If the chimps had learned the liquid/solid rule, and got the idea that a new lexigram was for a new item, they might learn more quickly. Indeed they did. Sherman and Austin were able to respond correctly the first time, or with only a few errors, instead of taking hundreds of trials as before. What had happened to produce this difference? What the animals had learned was not only a set of specific associations between lexigrams and objects or events. They had also learned a set of logical relationships between the lexigrams, relationships of exclusion and inclusion. More importantly, these lexigram-lexigram relationships formed a complete system in which each allowable or forbidden co-occurrence of lexigrams in the same string (and therefore each allowable or forbidden substitution of one lexigram for another) was defined. They had discovered that the relationship that a lexigram has to an object is a function of the relationship it has to other lexigrams, not just a function of the correlated appearance of both lexigram and object. This is the essence of a symbolic relationship.

The subordination of the indexical relationships between lexigrams (symbol tokens) and foods (referents or objects) to the system of indexical relationships between lexigrams is schematically depicted in three stages of development in Figure 3.3. Individual indexical associations are shown as single vertical arrows, mapping each token to a kind of object, because each of these relationships is independent of the others. In contrast, the token-token interrelationships (e.g., between lexigrams or words), shown as horizontal arrows interconnecting symbols, form a closed logical group of combinatorial possibilities. Every combination and exclusion relationship is unambiguously and categorically determined. The indexical reference of each symbol token to an object after symbolic reference is achieved is depicted with arrows reversed to indicate that these are now subordinate to the token-token associations.

In the minimalistic symbol system first learned by Sherman and Austin,
reference to objects is a collective function of relative position within this token-token reference system. No individual lexigram determines its own reference. Reference emerges from the hierarchic relationship between these two levels of indexicality, and by virtue of recognizing an abstract correspondence between the system of relationships between objects and the system of relationships between the lexigrams. In a sense, it is the recognition of an iconic relationship between the two systems of indices. Although indexical reference of tokens to objects is maintained in the transition to symbolic reference, it is no longer determined by or dependent on any physical correlation between token and object.

This makes a new kind of generalization possible: logical or categorical generalization, as opposed to stimulus generalization or learning set generalization. It is responsible for Sherman and Austin's ability to acquire new lexigrams and know their reference implicitly, without any trial-and-error learning. The system of lexigram-lexigram interrelationships is a source of implicit knowledge about how novel lexigrams must be incorporated into the system. Adding a new food lexigram, then, does not require the chimp to learn the correlative association of lexigram to object from scratch each time. The referential relationship is no longer solely (or mainly) a function of lexigram-food co-occurrence, but has become a function of the relationship that this new lexigram shares with the existing system of other lexigrams, and these offer a quite limited set of ways to integrate new items. The chimps succeed easily because they have shifted their search for associations from relationships among stimuli to relationships among lexigrams. A new food or drink lexigram must fit into a predetermined slot in this system of relationships. There are not more than a few possible alternatives to sample, and none requires assessing the probability of paired lexigram-food occurrence because lexigrams need no longer be treated as indices of food availability. Like words, the probability of co-occurrences may be quite low. The food lexigrams are in a real sense "nouns," and are defined by their potential combinatorial roles. Testing the chimps' ability to extrapolate to new lexigram-food relationships is a way of demonstrating whether or not they have learned this logical-categorical generalization, which is a crucial defining feature of symbolic reference.

At some point toward the end of the training, the whole set of explicitly presented indexical associations that the chimps had acquired was "re-coded" in their minds with respect to an implicit pattern of associations whose evidence was distributed across the whole set of trials. Did this re-coding happen as soon as they had learned the full set of combination/exclusion relationships among their lexigram set? I suspect not. Try to imagine yourself in their situation for a moment. You have just come to the point where you are not making errors. What is your strategy? Probably, you are struggling to remember what specific things worked and did not work, still at the level of one-by-one associations. The problem is, it is hard to remember all the details. What you need are aids to help organize what you know, because there are a lot of possibilities. But in the internal search for supports you discover that there is another source of redundancy and regularity that begins to appear, besides just the individual stimulus-response-reward regularities: the relationships between lexigrams! And these redundant patterns are far fewer than the messy set of dozens of individual associations that you are trying to keep track of. These regularities weren't apparent previously, because errors had obscured any underlying systematic relationship. But now that they are apparent, why not use them as added mnemonics to help simplify the memory load? Forced to repeat errorless trials over and over, Sherman and Austin didn't just learn the details well, they also became aware of something they couldn't have noticed otherwise, that there was a system behind it all. And they could use this new information, information about what they had already learned, to simplify greatly the mnemonic load created by the many individual rote associations. They could now afford to forget about individual correlations so long as they could keep track of them via the lexigram-lexigram rules.

What I am suggesting here is that the shift from associative predictions to symbolic predictions is initially a change in mnemonic strategy, a recoding. It is a way of offloading redundant details from working memory, by recognizing a higher-order regularity in the mess of associations, a trick that can accomplish the same task without having to hold all the details in mind. Unfortunately, nature seldom offers such nice neat logical systems that can help organize our associations. There are not many chances to use such strategies, so not much selection for this sort of process. We are forced to create artificial systems that have the appropriate properties. The crucial point is that when such a systematic set of tokens becomes available, it allows a shift in mnemonic strategy that results in a radical transformation in the mode of representation. What one knows in one way gets recoded in another way. It gets re-represented. We know the same associations, but we know them also in a different way. You might say we know them both from the bottom up, indexically, and from the top down, symbolically. And because this recoding is based on higher-order relationships, not the individual details, it often vastly simplifies the mnemonic problem and vastly augments the representational possibilities. Equally important is the vast amount of implicit knowledge it provides. Because the combinatorial rules
encode not objects but ways in which objects can be related, new symbols can immediately be incorporated and combined with others based on independent knowledge about what they symbolize.

The experimenters working with Sherman and Austin provided a further, and in some ways even more definitive, demonstration of the difference between indexical reference of lexigram-object correlations and symbolic reference in a subsequent experiment that compared the performance of the two symboling apes (Sherman and Austin) to a previous subject (Lana), who had been trained with the same lexigram system but not in the same systematic fashion. Lana had learned a much larger corpus of lexigram-object associations, though by simple paired associations. In this new experiment (see Figure 3.4), all three chimps were first tested on their ability to learn to sort food items together in one pan and tool items together in another (Lana learned in far fewer trials than Sherman and Austin). When all three chimps had learned this task, they were presented with new foods or tools to sort and were able to generalize from their prior behavior to sort these new items appropriately as well. This is essentially a test of stimulus generalization, and it is based on some rather abstract qualities of the test items (e.g., edibility). It shows that chimps have a sophisticated ability to conceptualize such abstract relationships irrespective of symbols. Of course, chimpanzees (as well as most other animal species) must be able to distinguish edible from inedible objects and treat each differently. Learning to sort them accordingly takes advantage of this preexisting categorical discrimination in a novel context. In this sense, then, what might be called an indexical concept of food and nonfood is not inherently related to the test items but stems from a prior learning experience.

This sorting task was followed by a second task in which the chimps were required to associate each of the previously distinguished food items with the same lexigram (glossed as “food” by the experimenters) and each of the tool items with another lexigram (“tool”). Initially, this task simply required the chimps to extend their prior associations with bins to two additional stimuli, the two lexigrams. Although all three chimps learned this task in a similar way, taking many hundreds of trials to make the transference, Sherman and Austin later spontaneously recoded this information in a way that Lana did not. This was demonstrated when, as in the prior task, novel food and novel tool items were introduced. Sherman and Austin found this to be a trivial addition and easily guessed without any additional learning which lexigram was appropriate. Lana not only failed to extend her categorization to the new items, the novelty and errors appeared to produce a kind of counter-evidence that caused her to abandon her prior training in a subsequent test. Though on the surface this task resembles the sorting task, these conflicting results demonstrate that there is a critical difference that underpinned the rote learning strategy used by Lana and favored the symbolic recoding used by Sherman and Austin. The difference is probably related to the fact that the sorting task involved a physical-spatial association of sign and object, whereas the lexigram “labeling” involved only temporal correspondence. Lana appeared not to be using these underlying qualities to
solve the task. For her, each lexigram object association was an independent datum, and so provided no information about other associations.

In contrast Sherman and Austin, as a result of their experience with a previous symbol system, recoded these new lexigram-object associations into two new symbolic categories that superseded the individual associations. It took them hundreds or thousands of trials to learn the first simple one-to-many associations. This was because they began with no systemic relationship in their still small lexigram repertoire for a general reference to “food” or “tool.” They had to learn them the hard way, so to speak, indexically. But as soon as they did learn these associations, they were primed to look for another higher-order logic, and once it was discovered, they were able to use this logic to generalize to new associations. Instead of hundreds or even thousands of trials, the availability of a symbolic coding allowed them to bypass further trials altogether, an incredible increase in learning efficiency. The chimps essentially knew something that they had never explicitly learned. They had gained a kind of implicit knowledge as a spontaneous byproduct of symbolic recoding.

I have chosen to recount this ape language study not because it portrays any particularly advanced abilities in chimpanzees, or because I think it is somehow representative. In fact (as noted earlier), more recent studies by these same experimenters, with a pygmy chimpanzee (or bonobo) named Kanzi, have demonstrated far more effortless and sophisticated symbolic abilities. Rather, I have focused on this earlier study because of the clarity with which it portrays the special nature of symbol learning, and because it clearly exemplifies the hierarchic relationship between symbolic and indexical reference. The *reductio ad absurdum* training ploy is particularly instructive, not because it is an essential element but because it provides an explicit constructive demonstration of the index-by-index basis of the eventual symbolic relationship. It also demonstrates how normal associative learning strategies can interfere with symbol learning. Indexical associations are necessary stepping stones to symbolic reference, but they must ultimately be superseded for symbolic reference to work.

**Unlearning an Insight**

The problem with symbol systems, then, is that there is both a lot of learning and unlearning that must take place before even a single symbolic relationship is available. Symbols cannot be acquired one at a time, the way other learned associations can, except after a reference symbol system is established. A logically complete system of relationships among the set of symbol tokens must be learned before the symbolic association between any one symbol token and an object can even be determined. The learning step occurs prior to recognizing the symbolic function, and this function only emerges from a system; it is not vested in any individual sign-object pairing. For this reason, it’s hard to get started. To learn a first symbolic relationship requires holding a lot of associations in mind at once while at the same time mentally sampling the potential combinatorial patterns hidden in their higher-order relationships. Even with a very small set of symbols the number of possible combinations is immense, and so sorting out which combinations work and which don’t requires sampling and remembering a large number of possibilities.

One of the most interesting features of the shift in learning strategy that symbolic reference depends upon is that it essentially takes no time; or rather, no more time than the process of perceptual recognition. Although the prior associations that will eventually be recoded into a symbolic system may take considerable time and effort to learn, the symbolic recoding of these relationships is not learned in the same way; it must instead be discovered or perceived, in some sense, by reflecting on what is already known. In other words, it is an implicit pattern that must be recognized in the relationship between the indexical associations. Recognition means linking the relationship of something new to something already known. The many interdependent associations that will ultimately provide the nodes in a matrix of symbol-symbol relationships must be in place in order for any one of them to refer symbolically, so they must each be learned prior to recognizing their symbolic associative functions. They must be learned as individual indexical referential relationships. The process of discovering the new symbolic association is a restructuring event, in which the previously learned associations are suddenly seen in a new light and must be reorganized with respect to one another. This reorganization requires mental effort to suppress one set of associative responses in favor of another derived from them. Discovering the superordinate symbolic relationship is not some added learning step, it is just noticing the system-level correspondences that are implicitly present between the token-token relationships and the object-object relationships that have been juxtaposed by indexical learning. What we might call a symbolic *insight* takes place the moment we let go of one associative strategy and grab hold of another higher-order one to guide our memory searches.

What I have described as the necessary cognitive steps to create symbolic reference would clearly be considered a species of “insight learning,” though my analysis suggests that the phrase is in one sense an oxymoron.
Psychologists and philosophers have long considered the ability to learn by insight to be an important characteristic of human intelligence. Animal behaviorists have also been fascinated with the question, Can other animals learn by insight? The famous Gestalt psychologist Wolfgang Köhler described experiments with chimpanzees in which to reach a fruit they had to "see" the problem in a new way. Köhler set his chimp the problem of retrieving a banana suspended from the roof of the cage and out of reach, given only a couple of wooden boxes that when stacked one upon the other could allow the banana to be reached. He found that these solutions were not intuitively obvious for a chimpanzee, who would often become frustrated and give up for a long period. During this time she would play with the boxes, often piling them up, climbing on them, and then knocking them down. At some point, however, the chimp eventually appeared to have recognized how this fit with the goal of getting at the banana, and would then quite purposefully maneuver the boxes into place and retrieve the prize. Once learned, the trick was remembered. Because of the role played by physical objects as mnemonic place-holders and the random undirected exploration of them, this is not perhaps the sort of insight that appears in cartoons as the turning on of a light bulb, nor is it what is popularly imagined to take place in the mind of an artist or scientist. On the other hand, what goes on "inside the head" during moments of human insight may simply be a more rapid covert version of the same, largely aimless object play. We recognize these as examples of insight solely because they involve a recoding of previously available but unlinked bits of information.

Most insight problems do not involve symbolic recoding, merely sensory recoding: "visualizing" the parts of a relationship in a new way. Transference of a learning set from one context to another is in this way also a kind of insight. Nevertheless, a propensity to search out new "perspectives" might be a significant advantage for discovering symbolic relationships. The shift in mnemonic strategy from indexical to symbolic use of food and food-delivery lexigrams required the chimps both to use the regularities of symbol-token combinations as the solution to correct performance, and to discover that features of the food objects and delivery events correspond to these lexigram combination regularities. In other words, they had to use these combination relationships to separate the abstract features of liquid and solid from their context of indexical associations with the food-delivery events. The symbolic reference that resulted depended on digging into these aspects of the interrelationships between things, as opposed to just mapping lexigrams to things themselves. By virtue of this, even the specific combinations of tokens cannot be seen as indexical, so that it is not just that the ability to combine tokens vastly multiplies referential possibilities, in the way that using two digits instead of one makes it possible to represent more numerical values. Which tokens can and cannot be combined and which can and cannot substitute for one another determines a new level of mapping to what linguists call "semantic features," such as the presence or absence of some property like "solidity." This is what allows a system of symbols to grow. New elements can be added, either by sharing reference with semantic features that the system already defines, or by identifying new features that somehow can be integrated with existing ones. Even separate symbol groups, independently constructed, can in this way become integrated with each other. Once the relationship between their semantic feature sets is recognized, their unification can in one insight create an enormous number of new combinatorial possibilities.

The insight-recoding problem becomes increasingly difficult as additional recoding steps become involved in establishing an association. For this reason, a child's initial discovery of the symbolic relationships underlying language is only the beginning of the demand on this type of learning/unlearning process. Each new level of symbols coding for other symbolic relationships (i.e., more abstract concepts) requires that we engage this process anew. This produces a pattern of learning that tends to exhibit more or less discrete stages. Since the number of combinatorial possibilities that must be sampled in order to discover the underlying symbolic logic increases geometrically with each additional level of recoding, it is almost always necessary to confine rote learning to one level at a time until the symbolic recoding becomes apparent before moving on to the next. This limitation is frustratingly familiar to every student who is forced to engage in seemingly endless rote learning before "getting" the underlying logic of some mathematical operation or scientific concept. It may also contribute to the crudely stagelike pattern of children's cognitive development, which the psychologist Jean Piaget initially noticed. However, this punctuated pattern of symbolic conceptual development is a reflection of symbolic information processing and not an intrinsic feature of developing brains and minds.

The ability of Sherman and Austin to discover the abstract symbolic references for "food" and "tool" provides an additional perspective on the difference between indexical associations and symbolic associations. Consider the potential conflict between the lexigram-object relationships they had previously acquired and this new set of associations. If their prior associations were supported only by the correlations in lexigram-object-reward occurrence, then re-pairing the same objects with a new lexigram would be
expected to partially if not totally extinguish the prior association. Although
it would be possible to provide additional contextual cues to enable the
chimps to decide which of two competing associative strategies to use (e.g.,
simply run trials without the alternatives available) and thus learn and ret-
tain both, there would still be interference effects (i.e., their prior associa-
tions might interfere both with relearning the new associations and with
shifting between them in different contexts). Unfortunately, data to assess
this are not available, but we can infer from Sherman and Austin’s learning
shifts, and their subsequent maintenance of the prior symbolic associa-
tions, that neither extinction nor interference was a significant problem.
Though it was not tested explicitly in this series of experiments, we should
expect that this should also distinguish Sherman and Austin from Lana. Cer-
tainly Lana’s rapid decline in performance when new items were added
points to such effects.

This ability to remember large numbers of potentially competing associa-
tions is an additional power of symbolic reference that derives from the
shift in mnemonic strategy to token-token relationships. Competition ef-
fects grow with increasing numbers of overlapping associative categories in
typical indexical reference relationships. Not only would the choice among
alternatives in any use become a source of confusion, but because they were
competing for reinforcement, each would weaken the association of the oth-
ers. Though some of the interference effects also attend symbol use, and
often are a cause of word retrieval errors and analysis delays, in terms of as-
sociative strength there is an opposite effect. Competing sets of overlapping
associative relationships on the indexical level translate into mutually sup-
portive higher-order semantic categories on the symbolic level. These be-
come sources of associative redundancy, each reinforcing the mnemonic
trace of the other. So, rather than weaken the strength of the association,
they actually reinforce it.

This helps to explain where the additional associative glue between words
and their referents comes from. Though token-object correlations are not
consistently available to the symbol user, indeed are rare, this loss of asso-
ciative support is more than compensated by the large number of other as-
sociations that are available through symbolically mediated token-token
relationships. Individually, these are comparatively weak associations, with
a low correlated occurrence of any two tokens in the same context; but they
are not just one-to-one associations. They are one-to-many and many-to-
one associations that weave symbol tokens together into a systematic net-
work of association relationships, and the pattern has a certain coded
isomorphism with relationships between objects and events in the world.
bolic and nonsymbolic associations as though they were the same sort of thing. Just as the contingencies of co-occurrence and exclusion in the same context determine the strengths of stimulus associations, so too do these statistics in language affect the strengths of word associations.

With each shift of referential control to a token-token system of relationships, it became possible for Sherman and Austin to add new lexical items to their growing symbol system with a minimum of associative learning, often without any trial-and-error testing. This produces a kind of threshold effect whereby prior associative learning strategies, characterized by an incremental narrowing of stimulus response features, are replaced by categorical guesses among a few alternatives. The result is a qualitative shift in performance. The probabilistic nature of the earlier stage is superseded by alternative testing that has a sort of all-or-none character. This change in behavior can thus be an indication of the subject’s shift in mnemonic strategy, and hence the transition from indexical to symbolic reference. The simplest indicator of this shift is probably the rate of acquisition of new lexical items, since this should be highly sensitive to the hundred- to thousand-fold reduction in trial-and-error learning required to reach 100 percent performance.

In young children’s learning of language, apparent threshold effects have long been noticed in vocabulary growth and sentence length. Vocabulary and utterance length are of course linked variables in two regards. First, the more words a child knows, the more there are to string together. But this does not simply translate into larger sentences. Creating a larger sentence in a human language cannot just be accomplished by stringing together more and more words. It requires the use of hierarchic grammatical relationships, as well as syntactic tricks for condensing and embedding kernel sentences in one another. Thus, not only does vocabulary need to grow, but the types of words must diversify. In other words, the regular discovery of new grammatical classes must be followed by a rapid filling of these classes with new alternative lexical items.

Each time a new logical group is discovered among a set of tokens, it essentially opens up one or more types of positional slots that can be filled from an open class of symbols. Each slot determines both a semantic and a grammatical category. Recall that although Sherman and Austin could add new food items to their lexigram “vocabulary” with little difficulty, when they had to learn to recode food items in terms of the higher-order semantic category “food,” they essentially had to start over. Their prior knowledge of the symbolic designations of distinct foods with respect to food-delivery modes was of no help. It may even have been a source of interference, since the same foods were now being linked with different lexigrams. But again, once this new symbolic association was established, adding new items proved trivial, usually involving no errors.

In the small symbol system initially learned by Sherman and Austin, the semantic features that were implicit in the few combinatorial possibilities available might be specified in terms of solid versus liquid and food versus delivery (of food). Discovering the combinatorial rules was the key to discovering these semantic features, and, conversely, these semantic features provided the basis for adding new symbols without needing to relearn new correlations. All that was necessary was prior knowledge of the object to be represented with respect to one or more of the relevant semantic features in order to know implicitly a token’s combinatorial possibilities and reference. Beginning with any initial core, the system can grow rapidly in repeated stages. Each stage represents a further symbolic transition that must begin with incremental indexical learning. But past experience at symbol building and a large system of features can progressively accelerate this process.

In summary, then, symbols cannot be understood as an unstructured collection of tokens that map to a collection of referents because symbols don’t just represent things in the world, they also represent each other. Because symbols do not directly refer to things in the world, but indirectly refer to them by virtue of referring to other symbols, they are implicitly combinatorial entities whose referential powers are derived by virtue of occupying determinate positions in an organized system of other symbols. Both their initial acquisition and their later use requires a combinatorial analysis. The structure of the whole system has a definite semantic topology that determines the ways symbols modify each other’s referential functions in different combinations. Because of this systematic relational basis of symbolic reference, no collection of signs can function symbolically unless the entire collection conforms to certain overall principles of organization. Symbolic reference emerges from a ground of nonsymbolic referential processes only because the indexical relationships between symbols are organized so as to form a logically closed group of mappings from symbol to symbol. This determinate character allows the higher-order system of associations to supplant the individual (indexical) referential support previously invested in each component symbol. This system of relationships between symbols determines a definite and distinctive topology that all operations involving those symbols must respect in order to retain referential power. The structure implicit in the symbol-symbol mapping is not present before symbolic reference, but comes into being and affects symbol com-
binations from the moment it is first constructed. The rules of combination that are implicit in this structure are discovered as novel combinations are progressively sampled. As a result, new rules may be discovered to be emergent requirements of encountering novel combinatorial problems, in much the same way as new mathematical laws are discovered to be implicit in novel manipulations of known operations.

Symbols do not, then, get accumulated into unstructured collections that can be arbitrarily shuffled into different combinations. The system of representational relationships, which develops between symbols as symbol systems grow, comprises an ever more complex matrix. In abstract terms, this is a kind of tangled hierarchic network of nodes and connections that defines a vast and constantly changing semantic space. Though semanticists and semiotic theorists have proposed various analogies to explain these underlying topological principles of semantic organization (such as +/- feature lists, dictionary analogies, encyclopedia analogies), we are far from a satisfactory account. Whatever the logic of this network of symbol-symbol relationships, it is inevitable that it will be reflected in the patterns of symbol-symbol combinations in communication.

Abstract theories of language, couched in terms of possible rules for combining unspecified tokens into strings, often implicitly assume that there is no constraint on theoretically possible combinatorial rule systems. Arbitrary strings of uninterpreted tokens have no reference and thus are unconstrained. But the symbolic use of tokens is constrained both by each token’s use and by the use of other tokens with respect to which it is defined. Strings of symbols used to communicate and to accomplish certain ends must inherit both the intrinsic constraints of symbol-symbol reference and the constraints imposed by external reference.

Some sort of regimented combinatorial organization is a logical necessity for any system of symbolic reference. Without an explicit syntactic framework and an implicit interpretive mapping, it is possible neither to produce unambiguous symbolic information nor to acquire symbols in the first place. Because symbolic reference is inherently systemic, there can be no symbolization without systematic relationships. Thus syntactic structure is an integral feature of symbolic reference, not something added and separate. It is the higher-order combinatorial logic, grammar, that maintains and regulates symbolic reference; but how a specific grammar is organized is not strongly restricted by this requirement. There need to be precise combinatorial rules, yet a vast number are possible that do not ever appear in natural languages. Many other factors must be taken into account in order to understand why only certain types of syntactic systems are actually em-

ployed in natural human languages and how we are able to learn the incredibly complicated rule systems that result.

So, before turning to the difficult problem of determining what it is about human brains that makes the symbolic recoding step so much easier for us than for the chimpanzees Sherman and Austin (and members of all other nonhuman species as well), it is instructive to reflect on the significance of this view of symbolization for theories of grammar and syntax. Not only does this analysis suggest that syntax and semantics are deeply interdependent facets of language—a view at odds with much current linguistic theory—it also forces us entirely to rethink current ideas about the nature of grammatical knowledge and how it comes to be acquired.