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 critter 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 association 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 referential 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 sustaining associations between a word and its reference, irrespective of their being correlated in experience, and those rote associations that are established 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.

CHAPTER THREE

Symbols Aren't Simple

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),[1] he argued that word meaning can be modeled by an element-by-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*.[2] 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 mind in terms of communication, essentially arguing that all forms of thought (ideas) are essentially communication (transmission of signs), organized by an underlying logic (or *semiotic*, as he called it) that is not fundamentally different for communication processes inside or outside of brains. If so, it might be possible to investigate the logic of thought processes 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 a 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 Cresent 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 hierarchy 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 laughter, 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.[3] 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 hierarchic 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.

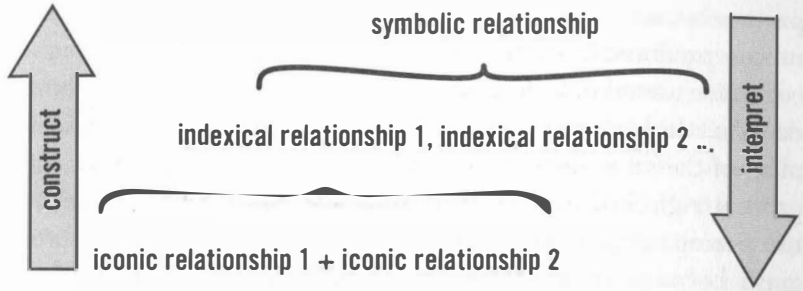


Figure 3.1 *The hierarchic relationships between the three fundamental forms of reference—iconic, indexical, and symbolic. Symbolic relationships are composed of indexical relationships between sets of indices and indexical relationships are composed of iconic relationships between sets of icons (diagrammed more pictorially in Figs. 3.2 and 3.3). This suggests a kind of semiotic reductionism in which more complex forms of representation are analyzable to simpler forms. In fact, this is essentially what occurs as forms are interpreted. Higher-order forms are decomposed into (replaced or represented by) lower-order forms. Inversely, to construct higher representation, one must operate on lower-order forms to replace them (represent them). In C. S. Peirce's terminology, each is an interpretive process, and the new signs substituted for the previous signs at a different level are "interpretants" of those prior signs (see text for details).*

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 reduceability (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.

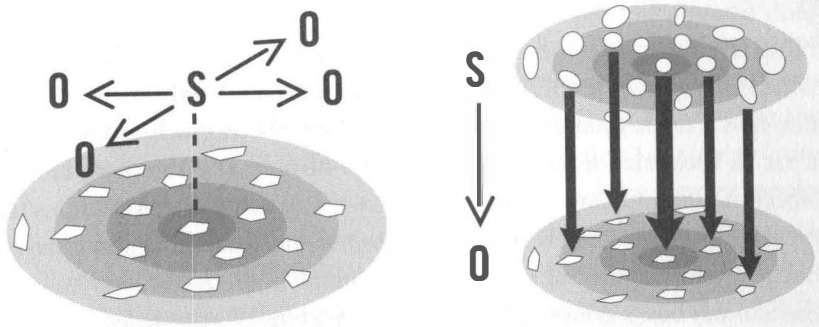
Consider the example of a symptom, like the smell of smoke. When I smell smoke, I begin to suspect that something is burning. How did my ability to treat this smell as an indication of fire arise? It likely arose by learning, because I had past experiences in which similar odors were traced to things 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 near. 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.



ICON: Associated by stimulus generalization or conventional similarity

INDEX: Associated by spatial-temporal correlation or part-whole contiguity

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*

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-

jects 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,[4] 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),[5] 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)[6] 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 chimp-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 lexigrams, the chimps were trained to chain lexigram pairs in a simple verb-noun relationship (a sequence glossed as meaning "give," which caused a dispenser to deliver a solid food, and "banana" to get a banana).[7] Initially there were only 2 "verb" lexigrams and 4 food or drink lexigrams to choose from, and each pair had to be separately taught. But after successful training of each pairing, the chimps were presented with all the options they had learned independently, and were required to choose which combination was most appropriate on the basis of food availability or preference. Curiously, the solution to this task was not implicit in their previous training. This was evident in the fact that some chimps tended stereotypically to repeat only the most recent single learned combination, whereas others chained together all options, irrespective of the intended meanings and what they knew about the situation. Thus they had learned the individual associations but failed to learn the system of relationships of which these correlations were a part. Although the logic of the combinatorial relationships between lexigrams was implicit in the particular combinations that the chimps learned, the converse exclusive relationships had not been learned. For example, they were not explicitly trained to avoid any number of inappropriate combinations such as "banana juice give." Though these errors are implicit for us, who treat them symbolically from the start, the combinatorial rules that allow pairing in some but not other cases was vastly underdetermined by the training experience (as it is also in a child's experience of others' word use).

It is not immediately obvious exactly how much exclusionary information is implicit, but it turns out to be quite a lot. Think about it from the naive chimpanzee perspective for a moment. Even with this ultra-simple symbol system of six lexigrams and a two-lexigram combinatorial grammar, the chimpanzee is faced with the possibility of sorting among 720 possible ordered sequences ($6*5*4*3*2*1$) or 64 possible ordered pairs. The training has offered only four prototype examples, in isolation. Though each chimp may begin with many guesses about what works, these are unlikely to be in the form of rules about classes of allowed and disallowed combinations, but rather about possible numbers of lexigrams that must be pressed, their positions on the board, their colors or shape cues that might be associated with a reward object, and so on. Recognizing this limitation, the experimenters embarked on a rather interesting course of training. They set out explicitly to train the chimps on which cues were not relevant and which combinations were not meaningful. This poses an interesting problem that every pet trainer has faced. You can't train what *not* to do unless the animal first produces the disallowed behavior. Only then can it be

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,

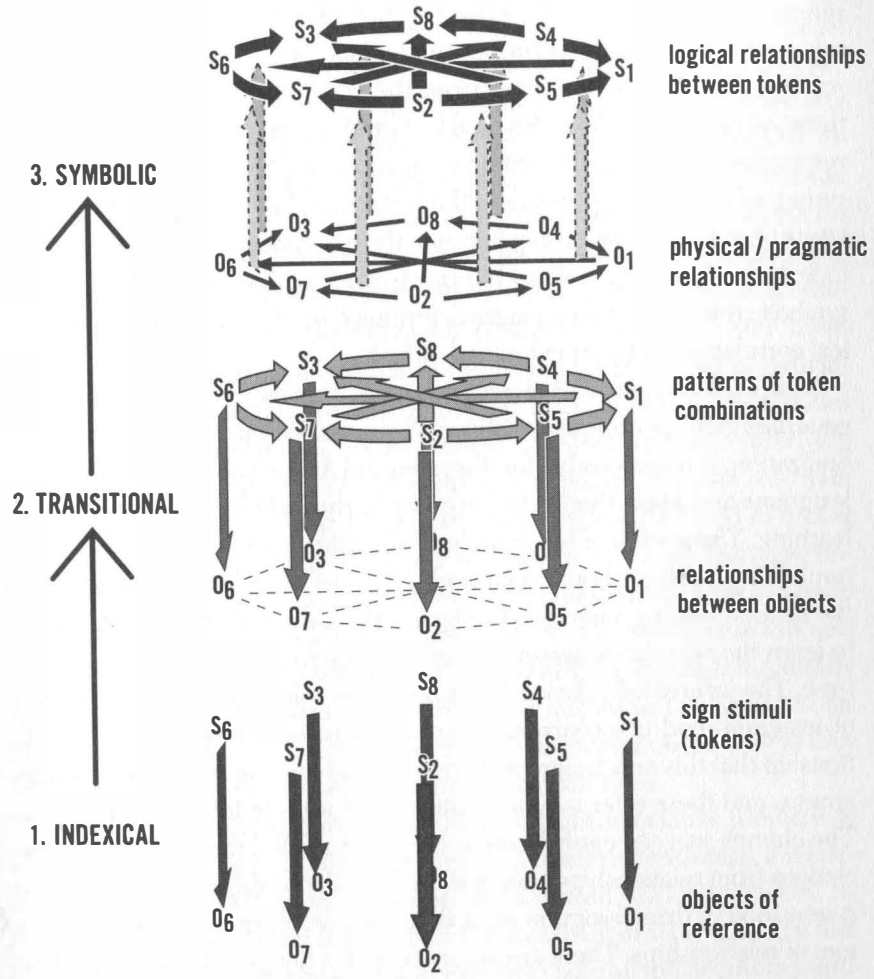


Figure 3.3 *A schematic depiction of the construction of symbolic referential relationships from indexical relationships. This figure builds on the logic depicted in Figure 3.2, but in this case the iconic relationships are only implied and the indexical relationships are condensed into single arrows. Three stages in the construction of symbolic relationships are shown from bottom to top. First, a collection of different indices are individually learned (varying strength indicated by darkness of arrows). Second, systematic relationships between index tokens (indexical stimuli) are recognized and learned as additional indices* (gray arrows linking indices). *Third, a shift* (reversal of indexical arrows) *in mnemonic strategy to rely on relationships between tokens* (darker arrows above) *to pick out objects indirectly via relationships between objects* (corresponding lower arrow system). *Individual indices can stand on their own in isolation, but symbols must be part of a closed group of transformations that links them in order to refer, otherwise they revert to indices.*

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 "recoded" in their minds with respect to an implicit pattern of associations whose evidence was distributed across the whole set of trials. Did this recoding 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 precedes the training. Each bin is eventually treated as indexical of this qualitative sensory and behavioral distinction, and so the ability to extend this association to new food and nonfood items involved stimulus generalization (though of an indirectly recognizable stimulus parameter).

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

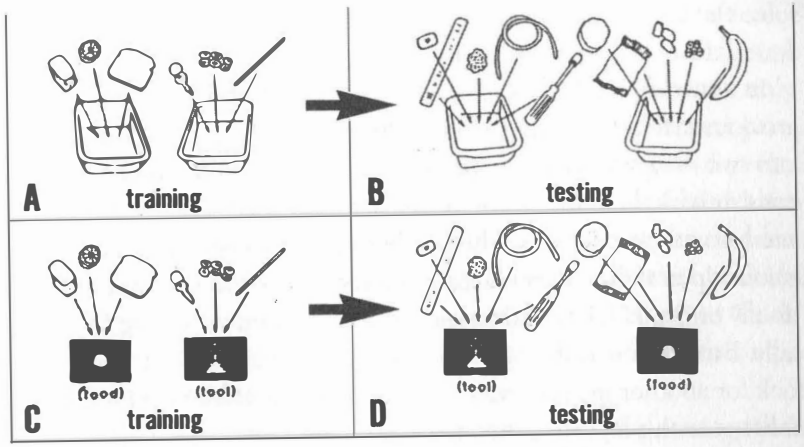


Figure 3.4 *Summary of part of a 1980 test of lexigram reference in chimpanzees by E. Sue Savage-Rumbaugh and her colleagues. This compares three levels of symbolic learning of lexigram reference by the chimps Sherman and Austin to the indexical learning of lexigram reference by another chimp, Lana, who is unable to complete tasks requiring symbolic reference. The panels on the left depict training trials and the panels on the right depict items added in test trials. Test trials introduced new lexigrams and tested to determine generalization to items for which there was no previous experience. The top task was merely a sorting task to determine that all animals understood the distinction between foods and tools (nonfood). The second task required identification with one of two lexigrams ("food," "tool"). Though all three learned it, only Sherman and Austin made the shift to symbolic categorization of reference and were able to generalize to new items (because of past symbol-learning experience). Lana was excluded from the remaining two procedures (not shown), where Sherman and Austin learned first to associate lexigrams to pictures of the foods and tools, and then to associate individual food and tool lexigrams with the appropriate general lexigram for food or tool.*

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 counterevidence 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 undermined 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.[5] 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 relationships 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.[9] 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.[10] 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 retain both, there would still be interference effects (i.e., their prior associations 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 associations, 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. Certainly Lana's rapid decline in performance when new items were added points to such effects.

This ability to remember large numbers of potentially competing associations is an additional power of symbolic reference that derives from the shift in mnemonic strategy to token-token relationships. Competition effects 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 others. 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 associative strength there is an opposite effect. Competing sets of overlapping associative relationships on the indexical level translate into mutually supportive higher-order semantic categories on the symbolic level. These become 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 associative support is more than compensated by the large number of other associations 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 network of association relationships, and the pattern has a certain coded isomorphism with relationships between objects and events in the world.

As a result of sharing many weak interpenetrating indexical links, each indexical association gains mnemonic support from a large number of others because they are multiply coded in memory. Together, their combined associative strengths make them far more resistant to extinction due to diminished external correlations with objects than are individual indexical associations. Thus, not only is symbolic reference a distributed relationship, so is its mnemonic support. This is why learning the symbolic reasons behind the bits of information we acquire by rote learning offers such a powerful aid to recall. How else could the many thousands of different words we use every day be retrieved so rapidly and effortlessly during the act of speaking or listening?

Numerous neuropsychological probes of semantic field effects demonstrate this for word meaning. Hearing, memorizing, or using a word can be a source of priming effects for subsequent recall or identification of other words in overlapping categories. For example, hearing the word "cat" might prime later memory tasks involving "dog" or "animal." Even more interesting is the fact that this also transfers to indexical associations involving these words as well. Receiving a mild electric shock every time you hear the word "cat" would cause you to learn to spontaneously produce physiological correlates of stress response (such as change in heart rate or galvanic skin response) upon hearing that word repeated. But a similar but less intense response will also be produced whenever you hear a word like "dog," even though there had never been shocks associated with these sounds. A lesser response will also be produced whenever you hear a word like "meow" or "animal," demonstrating lexical (word-word) associations, and in response to a rhyming word like "mat," demonstrating stimulus generalization effects. All of these distinct associative relationships are brought into relationship to one another by the symbolic relationship. Because each arouses an associative network that overlaps with that of the shock-conditioned word, the shared activation raises an arousal level also associated with shock. The extent of both the symbolic and indexical overlap appears to correlate with the extent of the transference. Though analogous to stimulus generalization, it is clearly different. There are no shared *stimulus* parameters that distinguish "dog" and "cat" from "car," which does not produce a similar priming. The difference is also reflected in the fact that there is an independent transference to words that rhyme, like "flat" or "sat." Rhyme associations are true stimulus generalization effects and also show some transference of physiological responses.

This analogy between effects involving shared stimulus features and shared semantic features shows that the brain stores and retrieves both sym-

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 employed 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.