Chapter 3
Word Learning and Theory of Mind

Learning a word is a social act. When children learn that rabbits eat carrots, they are learning something about the external world, but when they learn that *rabbit* refers to rabbits, they are learning an arbitrary convention shared by a community of speakers, an implicitly agreed-upon way of communicating. When children learn the meaning of a word, they are—whether they know it or not—learning something about the thoughts of other people.

What does this tell us about how words are learned? Maybe nothing. Just because the relationship between a word and its meaning is a social fact doesn't entail that one needs social competence or knowledge to learn this fact. After all, when dogs learn to obey the command "Sit!," they are also learning an arbitrary convention, one that exists in the minds of a community of English speakers. But dogs surely don't know this and can learn the command without ruminating about the thoughts of others. All they might do is associate the right behavior with the right sound, in the same way that they would learn other, nonsocial, facts. Maybe this is also true for how children learn words.

I argue here that it isn't. This chapter reviews evidence showing that children's word learning actually draws extensively on their understanding of the thoughts of others—on their theory of mind. Theory of mind underlies how children learn the entities to which words refer, intuit how words relate to one another, and understand how words can serve as communicative signs. I discuss certain alternatives, such as the view that word learning is done through general associative principles, aided by the careful naming practices of parents. I suggest that these fail to capture basic facts about language development.

This is a long chapter, but it is nowhere near long enough. It deals only with the most central ways in which theory of mind underlies word learning. But the role of theory of mind is something we return to several times in the chapters that follow, with regard to the learning of proper names and pronouns, the nature of artifact concepts, and the naming of representations. There is, however, much more to word learning than understanding the thoughts of other people, and this chapter concludes with a discussion of certain learning problems that must be solved in other ways.

The Associative Infant

A central aspect of word learning is figuring out which objects specific words refer to. This is far from all there is to word learning. Some words refer to nothing at all (words such as *of* and *the*), and others refer to things that are not objects (words such as *joke* and *number*). And even if we restrict ourselves to middle-sized objects such as rabbits and tables, we are stuck with Quine's problem, which is that children who hear a word and know that it refers to a rabbit are still faced with an indefinite number of possible meanings for this word: it could be a proper name, an adjective referring to the rabbit's color, and so on. But the ability to figure out which objects specific words refer to is nonetheless central to word learning; in its absence, some of these other issues don't even arise.

Associationism is a popular solution to this learning problem and has dominated psychological and philosophical thought for centuries. This is the view, defended in detail by empiricist philosophers such as John Stuart Mill, David Hume, and John Locke, that the mechanism for word learning is a sensitivity to covariation, one rooted in general principles of learning. If two thoughts occur at the same time, they become associated, and one gives rise to the other. Children learn the meaning of *rabbit,* then, because the word is used when they are observing or thinking about rabbits. As a result, the word and the thought become associated, and children could be said to have learned what the word means.

One version of this theory was adopted by B.F. Skinner (1957), who proposed that learning the meaning of a word is a matter of establishing—through reinforcement and punishment—a connection between a set of stimuli and a verbal response. Many computational models of word learning work in similar ways. Richards and Goldfarb (1986), for instance, propose that children come to know the meaning of the word *car* through repeatedly associating the verbal label ("car") with their experience at the time that the label is used. For those perceptual properties that repeatedly cooccur with the label, the association will strengthen, as with *four wheels;* for those that do not, as with *blue,* the association will weaken. As a result of this process, children come to associate the label "car" with those properties that only cars possess and could be said to have learned the meaning of the word.

Plunkett, Sinha, Møller, and Strandsby (1992) present a similar model, in which labels and images are fed through distinct "sensory" pathways into a network, and the network is trained to associate the two. Successful production occurs when the network generates the appropriate label in response to an image; successful comprehension occurs when the appropriate image is generated in response to a label. This is proposed as a theory of how young children learn words, and it is suggested that word learning—as well as language learning more generally (Plunkett, 1997)—is best explained through a connectionist architecture that is sensitive to statistical regularities in the environment.

Many facts about word learning are consistent with this perspective. Children's first words often refer to things they can see and touch, which is exactly what one would expect under an associationist learning procedure. And words are learned best in precisely the conditions in which an associative match would be easiest to make. If you want to teach someone *dog,* an excellent way to do so is to point to a dog, make sure the person is attending to it, and say "dog." If you wait until there are no dogs around and nobody is thinking about dogs, and then say "dog," the word will not be learned.

Lois Bloom summarizes one associationist theory of word learning as follows (1994, p. 221):

> Once the child learns something about objects and events, and about words *qua* words, word learning consists of good old-fashioned associative learning. In the beginning, the data for learning the meanings of language are in the circumstances of use in which children hear words and sentences. The meanings of early words like *cookie, gone, more,* and *mama,* or little sentences like "eat meat" or "throw ball" can be gotten from the words and their corresponding events. . . . Associative learning has now reappeared in contemporary theory as "connectionism." . . . Connectionism will continue to be debated in the realm of syntax for some time, but so far it offers a more parsimonious account of lexical learning than a theory based on a priori lexical principles.

This version of associationism posits more abilities on the part of children than the philosophical and computational versions discussed above, as Bloom implies that the input to the learning mechanism is already categorized in terms of objects and events, words and sentences. But what makes her view a bona fide associative theory is the proposal that the relationship between the words and what they refer to is established not through a process of reasoning and inference and not through specialized word-learning mechanisms, but through a sensitivity to covariation.

Because of this, Bloom's proposal retains one of the merits of associationism—parsimony. If she is right, word learning involves mental capacities that are present in the minds of other animals. Just as rats can come to associate a certain tone with a painful shock, children can learn to associate the word *cookie* with the sight and smell of cookies. Another virtue of this view is that it posits mechanisms that we know something about. There is no great mystery in how a brain could form associations between ideas or sensations that are present at the same time, and it is relatively straightforward to construct a computing device (either symbolic or connectionist) that does the same thing. Computational models of associative word learning (e.g., Gasser & Smith, 1998; Plunkett, Sinha, Møller & Strandsky, 1992; Richards & Goldfarb, 1986) are simple and elegant things.

But despite the merits of this proposal, it suffers from certain serious problems. One has to do with the input that children receive. Any associationist procedure requires that the right correlations are present in the environment. In the case of word learning, this entails that the words are presented at the same time that children are attending to what the words refer to. John Locke (1690/1964, p. 108) is clear about this: "For if we observe how children learn languages, we shall find that, to make them understand what the names of simple ideas or substances stand for, people ordinarily show them the thing whereof they would have them have the idea; and then repeat to them the name that stands for it: as *white, sweet, milk, sugar, cat, dog*."

But Locke is wrong. Words are not ordinarily used at the same time that their referents are being perceived. The best case for Locke is the learning of object names. But even for these, and even if we focus only on parent-child interactions within a supportive family environment, about 30 percent to 50 percent of the time that a word is used, young children are *not* attending to the object that the adult is talking about (Collins, 1977; Harris, Jones & Grant, 1983). Some of the time, for instance, that children hear "Want a cookie?," they will be staring at someone's face. But *cookie* doesn't mean face, and no child has thought that it does.

Solutions to the problem of noisy input can be found. It may be that some of the time *cookie* is used, for instance, children are not attending to cookies, but, in the fullness of time, the percepts that are most associated with the word are those elicited by encounters with cookies. So children who start off associating *cookie* with faces might, after hearing the word used over and over again, weaken this association and strengthen the association with cookies. An associative procedure doesn't need a perfect correlation, after all—just a statistically reliable one.

But this proposal makes the wrong prediction. It predicts that before children have enough data to converge on the right hypothesis, they should make frequent mapping errors, such as thinking that *cookie* means face. But this never happens. To account for this error-free learning, one might imagine that children are inherently cautious and use a word only when they have adequate statistical evidence for its meaning, such as hearing the word with a consistent referent a dozen times, across suitably different situations. But this also doesn't happen. Children do not wait: they can learn a word after hearing it used a few times in a single situation (e.g., Markson & Bloom, 1997; see chapter 2). The fact that object-name acquisition is typically both fast and errorless suggests that it is not a form of statistical learning.

Furthermore, Locke was assuming the Western model of adult-child interaction in which parents carefully name objects for their children. But this is not universal. In some cultures, this sort of ostensive labeling does not occur, and if children waited for adults to name objects that they were attending to, they would wait forever. Object names must instead be learned by attending to overheard speech (Lieven, 1994; Schieffelin, 1985). Despite this sort of cultural variation, all normal children learn the meanings of words.

Things get worse for an associationist account when one considers the problems that arise with the learning of names for things that you cannot see or touch. These include imaginary things, such as fictional characters, as well as abstract entities like numbers, geometrical forms, ideas, and mistakes. This is a problem that Locke and his contemporaries were well aware of, but it has never been solved. One might perhaps restrict the domain of associative learning to children's early vocabularies. But even these words can be surprisingly abstract. Nelson, Hampson, and Shaw (1993) examined the speech of 45 20-month-olds and found that only about half of children's nominals referred to basic-level object kinds; the rest referred to members of other conceptual categories, such as locations (*beach, kitchen*), actions (*kiss, nap*), social roles (*doctor, brother*), natural phenomena (*sky, rain*), and temporal entities (*morning, day*). Furthermore, despite their impoverished perceptual experience, blind children learn words, often at the same rate as sighted children (Landau & Gleitman, 1985).

The case against associationist theories of word learning gets stronger when we consider certain experiments that find that a statistical covariation between word and percept is neither necessary nor sufficient for word learning. And consider finally the fact (discussed below) that nonhuman primates, who are excellent at associative learning and have rich perceptual and motor systems, are quite abysmal at word learning.

I think the evidence is actually quite strong that associationism is simply false as a theory of early word learning. But to avoid confusion on this point, note that the objections above apply to associationism only under a particular, somewhat technical sense of the term. Another sense is highly general: to say that children "associate" a word with its meaning simply means that they have learned the meaning of the word. This is a harmless use of the term, but it has nothing to do with the empirical and testable proposal that word learning is done through a mechanism sensitive to statistical covariation, as proposed by Locke and others.

Furthermore, nobody is arguing against the view that children attend to the situation when they are learning a word. That would be crazy. No child has ever learned *dog* by hearing someone whisper the word in his ear as he lay in bed with his eyes closed. It may well be that children will learn *dog* best when they are attending to a dog when the word is used and that the more often they hear the word, the more likely they are to learn it. The issue is over why this is the case.

Finally, a rejection of associative theories of word learning does not entail the rejection of connectionism. There is a difference between the claim that word learning is not done through a sensitivity to statistical covariation and the stronger claim that the mechanisms that underlie word learning, whatever they are, do not emerge from connectionist learning algorithms. I suggest below that children learn the meanings of words through theory of mind. If this is right, then a *direct* connectionist implementation of word learning, in which sounds are associated with percepts, is unfeasible. (And this does preclude all connectionist theories of word learning that I am aware of.) But it leaves open the possibility that the mechanisms underlying word learning, while themselves not associationist, are somehow the product of associationist learning mechanisms. In particular, if a connectionist theory can account for the origin and nature of the relevant theory of mind capacities, then connectionism is consistent with the facts of early word learning. If it can't, it isn't.

The Augustinian Infant

How do children make the connection between words and what they refer to? One promising theory is that they do so through their understanding of the referential intentions of others. Instead of Locke, consider Augustine (398/1961, p. 11):

> When [my elders] named any thing, and as they spoke turned towards it, I saw and remembered that they called what they would point out by the name they uttered. And that they meant this thing and no other was plain from the motion of their body, the natural language, as it were, of all nations, expressed by the countenance, glances of the eye, gestures of the limbs, and tones of the voice, indicating the affections of the mind, as it pursues, possesses, rejects, or shuns. And thus by constantly hearing words, as they occurred in various sentences, I collected gradually for what they stood; and having broken in my mouth to these signs, I thereby gave utterance to my will.

To put this in more contemporary terms, children use their *naive psychology* or *theory of mind* to figure out what people are referring to when they use words. Word learning is a species of intentional inference or, as Simon Baron-Cohen (1995) has put it, mind reading.

This Augustinian perspective has not been popular in the last century. One of the central philosophical works of our time—Ludwig Wittgenstein's *Philosophical Investigations*—begins by ridiculing Augustine's assumption that children know about objects and people prior to language. His mistake, according to Wittgenstein, is the view that "the child could already *think,* only not yet speak." But Augustine's proposal is no longer seen as the goofy idea that it once was. Increasing evidence shows that some capacity to understand the minds of others may be present in babies before they begin to speak.

There are many names for this capacity, including *mind-reading, social cognition,* and *pragmatic understanding,* but in what follows, I use the term *theory of mind.* This is in part because I want to explore the implications that this proposal has for the study of language development in children with autism, which is often described as a deficit that particularly affects theory of mind (e.g., Baron-Cohen, Leslie & Frith, 1985). But two qualifications must be made about this usage. First, I'm using the term without any commitment to whether theory of mind is really a theory in any nontrivial sense (for different perspectives, see Baron-Cohen, 1995; Gopnik & Meltzoff, 1997; Leslie, 1994). And second, some researchers link the attainment of a theory of mind with the ability to pass the false-belief task (Wimmer & Perner, 1983), something that occurs at the age of three or four. There is considerable debate over why younger children do not pass this task and whether not passing it is really due to limitations in their understanding of the thoughts of others, but, in any case, I am using *theory of mind* in a broader sense that need not include the ability to reason about false beliefs.

What understanding do prelinguistic children have about the minds of others? Consider first sensitivity to what other people are attending to. By around nine months, a baby will naturally follow its mother's line of regard (Butterworth, 1991; Scaife & Bruner, 1975) and will also follow her pointing gestures (Murphy & Messer, 1977). At about the same age, babies can monitor their parents' emotional reactions to potentially dangerous situations and react accordingly. For instance, when seeing a spider, a baby will be less likely to approach it if its mother seems fearful than if she seems happy (Zarbatany & Lamb, 1985), and when babies are uncertain or hesitant, they check what their mother is looking at and how she is reacting (e.g., Bretherton, 1992).

These findings raise the question of what goes on when a baby follows the gaze of an adult. It might be that babies have an implicit assumption that the adult is attending to something and thinking about or reacting to that object. This would make gaze following a reflection of theory of mind (e.g., Baron-Cohen, 1995). Alternatively, gaze following might be the product of an automatic orienting procedure, either innate or learned, that is initiated by exposure to certain stimuli, such as eyes and faces, but has nothing to do with intentional attribution (e.g., Butterworth, 1991; Corkum & Moore, 1995; Perner, 1991).

One way to address this question is to ask what sort of stimuli elicit gaze following in babies. A study with 12-month-olds by Johnson, Slaughter, and Carey (in press) reports an intriguing finding. When exposed to a robot that interacts contingently with them, through beeping and light flashing, but that has no face, babies will nonetheless follow its "gaze" (the orientation of the front, reactive part of the robot), treating it as if it were a person. But they will not do so if a faceless robot fails to interact with them in a meaningful way. This suggests that gaze following is applied to entities that give some sign of having intentional states, regardless of their appearance, and supports the view that gaze following is related, at least for 12-months-olds, to intentional attribution.

Children are not merely passive observers of others. By about a year, they point on their own and then observe the adult's gaze, as if checking to see if they have succeeded in changing the adult's focus of attention (Bretherton, 1992). When they fail to capture an adult's attention in the right way, they often alternate between gazing at the object and at the adult and will modify their behavior until they succeed at getting the adult's attention (Golinkoff, 1986). Even at nine months, babies get adults to do things, such as open things, and play games—and they do so by first attracting the adult's attention and then making clear what they want, through gestures and vocalizations (Piaget, 1952). Finally, two-year-olds are sensitive to the knowledge of other people when they communicate. In one study, two-year-olds observed as an attractive toy was put on a high shelf. When later asking for help in retrieving the toy, they were more likely to name the toy and gesture to the location when their parent had not been present to witness the placement of toy, suggesting that even young children can take into account the knowledge and ignorance of other people (O'Neill, 1996).

All the examples so far involve children either trying to figure out the actions of an adult that they are interacting with or trying to manipulate the adult in some manner. But at least by their first birthday, children's abilities extend beyond this. They interpret abstract figures on a computer screen as goal-directed agents and expect them to behave in accord with canons of rational behavior (Gergely, Nádasdy, Csibra & Biró, 1995). They expect people to affect each other by action at a distance, but they have the opposite expectation about the behavior of inanimate objects (Spelke, Phillips & Woodward, 1995). They expect hands to move in goal-directed ways but do not have the same expectations about inanimate entities such as sticks (Woodward, 1998). One-year-olds can pretend that one object is another, as when pretending that a banana is a telephone, and by two—if not earlier—they can understand pretense by others. This appreciation of pretense shows up early in the words they use. If they know the word *telephone,* they have no hesitation, in the course of pretend play, in using this word to talk about a banana (Leslie, 1995).

These findings raise the possibility that children use these abilities to help them figure out what adults are intending to refer to when words are used. Again, this is not an alternative to the claim that children use perceptual information to learn words. After all, children are not telepathic: the only way they can infer the intentions of another person is by observing the properties of the situation, such as what the adult is looking or pointing at and what objects are present in the scene. The interesting question, then, is not whether children use such perceptual information (they plainly do) but what they do with it. Is it the basis for statistical reasoning, intentional inference, or both?

Both Locke and Augustine give the example of simple ostensive naming, in which an object is present and the child hears an adult name it. But what really goes on in this situation? In a fascinating series of studies, Dare Baldwin (1991, 1993b) tested babies in a context in which they were given one object to play with while another object was put into a bucket that was in front of the experimenter. When the baby was looking at the object in front of her, the experimenter looked at the object in the bucket and said a new word, such as "It's a modi!" This gives rise to a perfect Lockean correspondence between the new word and the object the baby was looking at. But 18-month-olds don't take *modi* as naming this object. Instead, they look at the experimenter and redirect their attention to what she is looking at, in this case, at the object in the bucket. And when later shown the two objects and asked to "find the modi," they assume that the word refers to the object the experimenter was looking at when she said the word—not the object that the child herself was looking at.

Similar results held even when the experiment was modified in the following way (Baldwin, 1993a). Two objects were hidden in different opaque containers. The experimenter opened one container, looked inside, said "It's a modi!," and then opened the other container, removed the toy, and gave it to the baby to look at and play with. After at least 10 seconds had passed, the experimenter removed the first object from the container and gave that to the baby as well. Again, when later tested, it was found that babies assumed that the word referred to the first object, the one that the experiment had named, despite the 10-second gap between hearing the name and seeing the object and despite the fact that they had interacted with another novel object in the meantime.

These studies show that a contiguity between word and percept is not necessary for word learning. Further work suggests that contiguity is also not sufficient. In another study, 15- to 20-month-olds were alone in a room with a novel object. When they looked at the object, they heard a disembodied voice (from a hidden adult outside the room) saying something such as "Dawnoo! There's a dawnoo!" Under this circumstances, they did not learn the word (Baldwin et al., 1996). That is, even with a perfect association between hearing a word and attending to an object, young children will make the connection only if they have some warrant to believe that it is an act of naming—and for this, the speaker has to be present. (Adults, of course, could learn the word in the above situation, presumably not because we are more associationist than children but because we would infer that the disembodied voice is actually an act of naming by a person who we can't see.)

Michael Tomasello and his colleagues have found that older children show a more subtle appreciation of intentional cues. In one study with 24-month-olds (Tomasello & Barton, 1994), the experimenter looked into the child's eyes and said, "Let's find the toma. Where's the toma?" Both the experimenter and the child then approached a row of five buckets, each of which contained a novel object. In the "without search" condition, the experimenter immediately withdrew an object from one of the buckets, held it up with an excited look, gasped "Ah!," and then handed it to the child. In the "with search" condition, the experimenter withdrew one object, scowled, and put it back; did the same with a second object; and then withdrew a third object, held it up, looked excited, gasped "Ah!," and handed it to the child. After both conditions, the experimenter then extracted the remaining objects from each of the buckets, saying each time, "Let's see what's in *here?*" When later shown the five objects and asked to find the toma, children performed equally well in both conditions, picking out the object that the experimenter seemed happy with, despite the fact that it was not the last object they saw and, in the "with search" condition, not the first.

Success in this task could not be due to any procedure based on direction of gaze. It instead had to result from children's sensitivity to what the adult's goal was and when it was satisfied, as indicated by cues such as the experimenter's expressions of happiness and the fact that the object was given to the child. Modified versions of these studies have found the same abilities with 18-month-olds (Tomasello, Strosberg & Akhtar, 1996).

Another set of studies explored the acquisition of verbs. The experimenter would introduce a new verb, saying, "I am going to plonk Big Bird!" Then she would perform an action and say "There!" and then perform another action and say "Whoops!" This was meant to give the impression that the first action was intentional and the second was accidental. (In another condition, the order of the intentional and accidental actions was reversed.) Two-year-olds were sensitive to this emotional cue of intent: When later asked "Can you go plonk Big Bird?," they tended to imitate the intentional action, not the accidental one (Tomasello & Barton, 1994).

Lexical Contrast

The Phenomena

Consider children who hear a new word in a situation in which intentional factors such as direction of gaze suggest that a certain object is the referent of the word. But the object already has a name. For instance, children might hear a word, note that the person is looking at several objects, including a rabbit—but already know the word *rabbit*. What do children do in such a situation?

Abundant evidence suggests they are biased to think that the word does not have the same meaning as *rabbit*. The original study to test this was done by Kagan (1981), as part of large-scale study of development in the second year. Children were shown three objects, two of them familiar (a doll and a dog) and one unfamiliar. Children were allowed to play with the three objects and were then asked to "Give me the zoob." By the age of about 22 months, both American and Fijian children tended to choose the novel object, suggesting that they believe that a novel word does not refer to objects that already have names.

A concern with this study, however, is that these children were repeatedly tested over several months, which might have helped train them into making the appropriate response. A later study by Markman and Wachtel (1988) is immune from this concern. Preschool children were shown two objects, one familiar and one novel, such as a banana and a whisk, and were presented with a new word, as in "Show me the fendle." They tend to interpret the new word as naming the whisk, not the banana. If only the banana is present, children are prone to take a novel name as referring to a part of the object, not the object itself. Markman and Wachtel (1988) explain this in terms of a mutual exclusivity principle, one that biases children to think that words should not have overlapping reference or, equivalently, that each object can have only one label.

These findings have been replicated and extended in many ways (e.g., Au & Glusman, 1990; Hall, 1991; Hutchinson, 1986; Merriman & Bowman, 1989). A similar effect shows up with verbs. Golinkoff, Shuff-Bailey, Olguin, and Ruan (1995) and Merriman, Marazita, and Jarvis (1995) found that if you show children two actions, one familiar and one unfamiliar, and produce a novel verb (as in "Where's gorping?"), they will tend to assume that this new verb refers to the action they do not already have a name for.

In a study by Markman and Wasow (reported in Woodward & Markman, 1997), even 15-months-olds appear to be sensitive to mutual exclusivity. When shown an object with an already known name, such as a spoon, and asked "Can you show me the modi? Where's the modi?," they tend to look around for a referent for the word. If there is a bucket present, they will look in the bucket, searching for the modi, reflecting a tacit expectation that the spoon is not likely to be the modi.

Mutual exclusivity can also explain certain facts about language development. A child who calls cats "dogs" might stop doing so once learning the word *cat,* and neologisms, such as *climber,* drop off once children learn the correct English word for the object, in this case, *ladder* (Clark, 1987). Or consider children's problems with superordinates. Macnamara (1982) reports that his two-year-old son Kieran rigorously refused to call a single dog "an animal," and the same phenomenon has been found experimentally (Callanan & Markman, 1982; Macnamara, 1982, 1986). Perhaps children are reluctant to use superordinates (such as *animal*) because they already have names (such as *dog*) for the objects that the superordinates refer to (Markman, 1989).

Mutual exclusivity applies only relative to a particular language. If a Spanish-English bilingual child knows the name of something in

English and then is given a second name for the same object in Spanish, she has no problem learning it (Au & Glusman, 1990). Furthermore, mutual exclusivity is a bias or default assumption, not an immutable restriction (Markman, 1992; see Behrend, 1990, and Nelson, 1988, for discussion). After all, languages frequently violate mutual exclusivity. Consider *dog, puppy, pet, animal,* and *Fido,* all of which could label the very same object. Words that have overlapping reference can be learned by children; it is just that they are harder to learn.

Mutual exclusivity is not a subtle phenomenon: you don't need to test dozens of children in careful laboratory conditions to see it at work. I often sit with my son Max and look at picture books with him, and like a typical Western parent, I point to pictures and say "What's that?" Sometimes he gets it wrong: when looking through a book about vehicles, for instance, he might call a dump truck "a tractor." Since I try to be supportive, I don't tell him he's wrong. Instead, I point to the same picture and say cheerfully "That's a dump truck!" This has the same effect, however, as a direct correction. By about 22 months he would look up at me and say, with considerable seriousness: "Dump truck. *Not* tractor. Dump truck."

Its Nature and Origin

This bias could be a specifically lexical phenomenon, a fact about how words work that is either innate or acquired in the course of language development (e.g., Mervis, Golinkoff & Bertrand, 1994). Or it could be a special case of a general principle of learning, one guiding children to prefer one-to-one mappings as part of a general tendency to exaggerate regularities (e.g., Markman, 1992). A third possibility, which I explore here, is that mutual exclusivity is a product of children's theory of mind.

This view has been defended by Eve Clark (1987, 1993, 1997; see also Gathercole, 1987). The main idea is that children's bias against lexical overlap can be explained in terms of a pragmatic principle—the principle of contrast—which states that every difference in form corresponds to some difference in meaning. There are no synonyms.

This is a controversial claim, and whether you find it convincing will depend on what you think meaning is. If you equate meaning with reference, for instance, then there certainly are synonyms, such as *cop* and *policeman,* that pick out the very same entities in the world. But Clark is endorsing a theory of meaning that includes a host of other factors as well, including considerations of register (*cop* versus *policeman*), emotive qualities (*statesman* versus *politician*), and dialect differences (*tap* versus *faucet*), and, under this very fine-grained notion of meaning, it is plausible that such a principle applies.

As Clark notes, the principle of contrast is an old idea and has been proposed, in slightly different variants, by Bloomfield, de Saussure, and von Humboldt. It is typically seen as resulting from the psychologies of individual people: speakers will not use terms interchangeably. Should synonymous terms somehow come to exist—through language contact, for instance—children would not learn them.

Why not? The answer might lie in children's inferences about the thoughts of others. Consider again the Markman and Wachtel (1988) study in which children were shown a banana and a whisk and asked to "Show me the fendle." A child might reason as follows (implicitly, of course):

1. I know that a banana is called *banana.*
2. If the speaker meant to refer to the banana, she would have asked me to show her the banana.
3. But she didn't; she used a strange word, *fendle.*
4. So she must intend to refer to something other than the banana.
5. A plausible candidate is the whisk.
6. *Fendle* must refer to the whisk.

Statements 1 and 2 capture what Clark calls the principle of conventionality—that words have fixed conventional meanings; 4 captures the principle of contrast; and 5 is the result of the child's assumption that new words presented in a neutral context such as "Show me the" are names for basic-level object kinds and not colors, parts, superordinate kinds, and so on (see chapters 4 and 6). If—contrary to how the mind actually works—children were prone to take a new noun used in a neutral context as a color term, then the color of the banana would be salient. But since children are biased to take a new word in this circumstance as describing a basic-level object kind, and since another object is present, then the principle of contrast explains why children are drawn to the whisk as the referent of *fendle.*

This pattern of reasoning applies to any communication system, not just words. To see this, imagine the following situation. A cube and a sphere are in front of you. An experimenter holds up a red card and motions to you to hand over the cube. You obey, and the experimenter thanks you. Game over. Now you play again. The experimenter holds up the red card, and you hand over the cube. (If you hand her the sphere, she shouts "Wrong!") And again and again.

All of a sudden she holds up an orange card. What do you do? It seems likely that you would go through the same reasoning as above:

1. I know that the red card means "hand over the cube."
2. If the experimenter had meant for me to hand over the cube, she would have held up the red card.

3. But she didn't; she used a different card, an orange one.
4. So she must intend me to do something other than hand over the cube.
5. A plausible candidate is to hand over the sphere.
6. The orange card must mean "hand over the sphere."

The idea here is that a mutual-exclusivity bias will arise in the learning of any conventional communicative system. It is not limited to language. But it is not entirely broad either; it does not apply to the learning of any system of mappings. Nothing should stop children, say, from assuming that a single object has many properties, so long as the properties are not inherently conflicting. To see this, consider children who see a banana and know two things about it: it is called *banana,* and it is yellow. They will be loath to accept that it could have another name but should not have the same unwillingness to learn that it could have another *property,* such as being tasty.

We can therefore distinguish these theories in terms of their scope. A strictly lexical theory predicts that contrast should apply only to words, a simplicity-of-mapping theory predicts that it should apply to all domains, and the theory-of-mind proposal predicts that it should apply only to communicative situations.

Little research is available on this topic. It would be nice to know, for instance, if mutual exclusivity applies when children learn new gestures or when they learn the sounds that different animals make. One set of studies, however, by Gil Diesendruck and Lori Markson (under review), does bear directly on this issue.

In one experiment, three-year-olds were presented with two unfamiliar objects and told a novel name for one of them ("This is a mep"). When exposed to both objects and asked about the meaning of a second different name ("Can you show me a jop?"), they tended to think that it referred to the other unnamed object, replicating previous studies. In another condition, a different group of children was shown two objects, told a novel *fact* about one of them ("My sister gave this to me"), and then asked to select the referent of a different fact ("Can you show me the one that dogs like to play with?"). The same result ensued as in the word study: children tended to choose the other object as the referent of the new information. This is presumably because they were reasoning that if the experimenter had intended to refer to the first object, she would have referred to it by using the original fact ("the one my sister gave me"); she would not have introduced a different fact.

Diesendruck and Markson went on to test a further prediction—that if children are using pragmatic reasoning about the adult's intentions in using the new fact, they should be less inclined to produce such a response in a two-speaker scenario, where the second speaker lacks mutual knowledge with the child. That is, if one speaker tells the child "My sister gave this to me" about one object, and then a different speaker, new to the discourse context, enters the room and asks "Can you give me the one that dogs like to play with?," the prediction is that children should now choose each of the objects with equal frequency. This is precisely what occurred.

Taken together, these findings support the notion that lexical contrast has its origin in children's expectations about the communicative behavior of others. It applies just as strongly when children are taught facts about objects as when they are taught words, and it does not apply when the pragmatic expectations are modified, as in the two-speaker condition. In sum, the reasoning that underlies the assumption of lexical contrast is not limited to words or to a general bias in favor of one-to-one mappings.

An important difference between words and facts, however, was discovered when Diesendruck and Markson did the two-speaker condition with novel words. In this condition, one speaker tells the child "This is a mep," and a different speaker enters the room and asks "Can you show me the jop?" Here children chose the object that wasn't originally labeled as the *mep*, the same as they did in the one-speaker condition with novel words—but different from their behavior in the two-speaker fact condition.

This suggests that children know something about words that isn't true about facts. Words have public meanings. If one person says of an object, "My sister gave this to me," there is no reason to expect this utterance to relate to the linguistic behavior of someone who later arrives on the scene. But if one person describes an object as a *mep*, it would be reasonable for a child to infer that other people know this word as well. Hence when a second person, new to the discourse content, asks for the jop, children could infer that if she meant to refer to the mep, she would have asked for it. Since she didn't, she must have meant to refer to the second object. This analysis raises the question of how children come to understand that words have this special property of having public meanings, something I return to later in the chapter.

Objections

One argument against this attempt to reduce lexical contrast to children's assumptions about the communicative goals of other people goes as follows. If lexical contrast applies even in cases where the speaker's intent to use the word to refer to a given object is entirely clear, this would suggest that the bias has a nonintentional origin.

Woodward and Markman (1997) report an experiment by Nowinsky and Markman that involves a familiar object, such as a shoe, and an unfamiliar one, such as a whisk. In one condition, children were asked "Please hand me the item," and they tended to hand over the whisk. Since *item* is an unfamiliar word, this replicates the standard mutual exclusivity finding. But in another condition, the experimenter would first point to the shoe and say "Look at the item," making her intent to refer to the shoe perfectly clear, *and then* would show the children both objects and ask "Please hand me the item." Children still treat *item* as referring to the whisk. This finding led Woodward and Markman to conclude that mutual exclusivity cannot reduce entirely to a pragmatic bias since, after all, there is abundant pragmatic evidence that *item* refers to the shoe.

Is this evidence against a pragmatic version of mutual exclusivity? The problem with this conclusion is that in the second condition, the referential intent *isn't* clear; it is conflicting (see Clark, 1997, pp. 35–36, for discussion of a similar case). On the one hand, the speaker has plainly just called the shoe "an item." On the other hand, the principle of contrast states that *item* cannot mean "shoe," since *shoe* means "shoe." If children possess the principle of contrast, they are in a difficult situation. They might choose to entirely ignore the fact that the shoe was called "the item." Or they might treat *item* as a superordinate term that could refer to either the shoe or the whisk (which is in fact the correct interpretation of the word). Either interpretation is consistent with the fact that children end up choosing the whisk as the item, and so this finding is not evidence against the theory of mind analysis.

A different sort of objection, raised by Dan Sperber (1997), applies to the more general program of explaining early word learning in terms of theory of mind. Complicated reasoning about the thoughts of other people is slow and difficult. Even if one accepts that children have the requisite background knowledge for these inferences, is it really reasonable to believe that they can carry them out in the course of word learning?

It is true that the sort of inferences involved in early word learning probably *are* slow and difficult for young children. But this is actually consistent with the developmental facts. For instance, 18-month-olds can cope with discrepant-looking situations, but younger children cannot. Baldwin (1991) finds that 16- to 17-month-olds succeed at learning words in a joint-attention condition—that is, when the child and adult are both looking at the same object—but do not learn the word in a discrepant condition, when the child and the adult are looking at different objects. This suggests that 16- to 17-month-olds understand the relevance of attentional focus (if they didn't, the children would have simply taken the new word as naming what they were looking at when they heard it), but that the processing demands of this task were too much for them. Similarly, while one- and two-year-olds show a bias against lexical overlap, it is initially quite weak; the strength of this bias grows in the years that follow (see Merriman, Marazita & Jarvis, 1995, for review). If one does something over and over again, a slow and effortful process can become fast and easy, and this might be what happens for mutual exclusivity in particular and (as argued in chapter 2) for word learning in general.

After all, the use of theory of mind is a ubiquitous part of communication. Understanding a sentence involves more than using lexical and syntactic knowledge to decode a message. It is an act of intentional interpretation, involving a mutual expectation of cooperation between speaker and listener (see Grice, 1975; Sperber & Wilson, 1986). This explains the resolution of ambiguity, as well our understanding of nonliteral language such as metaphor, irony, humor, sarcasm, and politeness. If someone asks you "Would you mind telling me what time it is?," it would be perverse or rude to focus on the literal form of the question and answer "No, I wouldn't mind at all." Nonliteral language is processed very rapidly by adults (Gibbs, 1983) and comes naturally to young children, who find it easier to attend to the intended meaning of a sentence than to its literal form (Beal & Flavell, 1984). It is clear, then, that at some point, relatively early in development, such inferences do become second nature.

Its Function

What is the assumption of lexical contrast good for? What would happen to a child who was normal in every way except that she had no bias to assume that people use different words to convey different meanings? While a normal child who hears words such as *Fido, dog, white, animal, pet,* and *tail* applied to a dog would assume that these words all have different meanings, our imaginary child will make no such assumption and could take them all as mutually synonymous. It is not obvious that this child will be at any long-term disadvantage. Since these words really do refer to different things, sooner or later the child will converge on their right meanings. She will notice, for instance, that cats are also called "animals" and infer that *animal* and *dog* are not synonymous. Syntactic cues could aid in distinguishing proper names, common nouns, and adjectives, and pragmatic and contextual cues could distinguish names for parts and wholes.

But this child will face a subtler problem (pointed out to me by Gregory Murphy), which does require a mutual exclusivity bias. Many of the words that children learn are going to refer to categories whose boundaries are arbitrary, at least from the child's standpoint. A bat is not a bird, a bean-bag chair is not a pillow, boots are not shoes, and so on. An avoidance of lexical overlap might be essential to learning about the status of entities that lie on these boundaries and hence could rescue children from overextensions.

Consider the plight of the child who comes to the reasonable, but wrong, view that *ring* refers to any piece of jewelry that encompasses the hand or wrist and so includes bracelets. What can correct this impression? If children call a bracelet "a ring," they might be corrected. But it is implausible that this sort of error and subsequent correction are necessary and that if they never occurred, the misinterpretation would never go away. (If error and correction were necessary, then a significant number of people reading this will be surprised to hear that a bracelet is not a ring.) Instead, children who possess the assumption of lexical contrast can learn that bracelets are not rings just by hearing them called "bracelets." And the same for learning that bats are not birds and bean-bag chairs are not pillows. The precise boundaries of categories can be acquired though noting how words contrast with another.

Note, however, that it is not *just* the fact that the bracelet is called "bracelet" that leads children to figure out that it cannot be a ring. After all, when a child hears a poodle described as "an animal," "a pet," or "a poodle," she does not conclude that poodles actually aren't dogs. As Clark (1998) puts it, children appreciate that people can take multiple perspectives when describing an entity; only when two words are understood as involving the same perspective does the principle of contrast (rooted in theory of mind) kick in, and the child infers that one of the words has to go. In this case, children know that both *bracelet* and *ring* are basic-level object names (see chapters 4 and 6), and this leads them to the insight that they have overextended *ring*—that rings are one kind of thing and bracelets are another.

This sort of account can be taken too far. Ferdinand de Saussure (1916/1959) thought names were entirely acquired and understood through this sensitivity to opposition. But this cannot be right. Children start off generalizing words in a constrained way, and even their earliest words tend to be used in a roughly appropriate manner. Although children might think that a bracelet could be called "a ring," they will not extend *ring* to refer to an aardvark, regardless of whether they know the word *aardvark.* This pragmatic understanding of how words relate to one another is instrumental in fine-tuning children's understanding of words; it is not necessary for word learning in general.

The Origin of Words

The discussion above assumes that children already have some implicit notion of what a word is. To see why, consider a child who sees her father point to a dog and say "dog." Suppose the child can infer that her father is referring to a dog. Still, all she could really be said to know is that he intended to refer to the dog and produced a sound as he did so. More is required for the child to know that *dog* is a word that refers to dogs. She has to infer that this sound is related to the act of reference. And she has to realize that anyone, including herself, can refer to dogs by making the same sound. In other words, the child has to make the inferential leap from hearing someone say "dog" when referring to something to the conclusion that *dog* is a word, a Saussurian sign.

Words have unusual properties. Saussure famously stressed their arbitrary nature, the fact that it is an accident of history that a particular form gets mapped onto a particular meaning. But he also noted that words are bidirectional with regard to comprehension and production. James Hurford (1989) points out that this fact is not a logical necessity; there are other ways a communication system could work and other expectations children might have. They could infer that the sound is a symbol in an asymmetric communication system.

A lot of communication is asymmetric. Suppose a child observes her father react to a wasp by gasping. It would be mistaken for her to assume that if on another occasion *she* gasped, her father would think there was a wasp present. Some dogs come to their owner when they are called, but no dogs make the inference that if they were to produce the same sound, their owner will obediently run to them. My computer sometimes asks me questions ("Are you sure you want to permanently remove these items?"), but if I were to type in the same questions, no communication would take place. These are all non-Saussurian systems.

Hurford notes that humans could conceivably have evolved to use non-Saussurian languages. Here is how such a language could work. Imagine two speakers (A and B) and two things to talk about (dogs and cats). If Speaker A wants Speaker B to think about dogs, she says "dog," and if she wants him to think about cats, she says "cat." In contrast, if Speaker B wants Speaker A to think about dogs, he says "chien," and if he wants her to think about cats, he says "chat."

This system is non-Saussurian: while speaker A produces "dog" to communicate about dogs, she might not understand the word "dog" if she heard it. Such systems are sometimes used in the real world. Two bilinguals could converse, each of them speaking in his or her preferred language. And a blind person and a deaf person could communicate this way: the blind person could transmit in a sign language such as ASL and receive in a spoken language such as English, and the deaf person could transmit in English and receive in ASL. So why haven't such systems evolved?

Hurford argues that while non-Saussurian systems might be usable, they are almost impossible to learn. With two speakers, it is bad enough; each has to generate a language and teach it to the other. As the number of speakers increases, the burden increases exponentially. A child who is born into a community of 10 speakers has to not only learn 10 languages but also teach her own language to each of the 10 other speakers. Using a computer simulation, Hurford found that an animal that adopts a Saussurian strategy (and learns words as bidirectional symbols) communicates better than an animal who use a non-Saussurian communication system (see Skyrms, 1996, for a similar analysis).

Under this view, the Saussurian nature of words has evolved through natural selection as part of the evolution of language. But there are reasons to favor an alternative, which is that children's assumption that words will be Saussurian signs is a natural consequence of their theory of mind (see Lewis, 1969). This fact about words is not something that has evolved once in the history of our species; it is discovered anew by every child.

One argument for this alternative is that this Saussurian assumption of bidirectionality is not limited to language. A similar phenomenon shows up in other domains. In an experiment by Meltzoff (1988), for instance, 14-month-olds are shown an unusual act that achieves a goal, as when an adult bends at the waist to touch a panel with her forehead, causing a light to go on. When shown this, babies will often spontaneously imitate the act. And when 18-month-olds are shown an action that an adult tries to do and fails, such as attempting to hang a loop on a metal prong, they will often imitate the entire successful action, even though they had never seen it before (Meltzoff, 1995).

Actions are not symbols, but what goes on in Meltzoff's experiments might be very similar to what happens when children learn a new word. To see this, consider the inference that babies made in the first study:

Scene:	An adult touches a panel with his or her forehead, and a light goes on.
Goal:	Turning on the light
Action:	Touching the panel with the forehead
Inference:	The light can be turned on by touching the panel with one's forehead.

The second study suggests that the same sort of inference can be made when the action and the goal are themselves inferred and not actually witnessed:

Scene:	An adult tries to hang a flexible loop on a metal prong and fails.
Goal:	Hanging the loop on the prong
Action:	Moving the loop to the appropriate location
Inference:	The loop can be hung on the prong by moving it to the appropriate location.

Evidence suggests that children's understanding of the goal in the above studies is based on inference about the intentions of the actor and not on observation of the physical motions. When a separate group of babies in the Meltzoff (1995) study saw the same physical motions performed by a mechanical handlike device, they did *not* imitate the attempted actions; they do so only if they see an intentional agent do it. Similarly, when habituated to a hand reaching toward and grasping one of two objects, six-month-olds looked longer when the hand subsequently made the same reaching motion to grasp the second object than when the hand made a physically different reaching motion to grasp the original object. But they showed the reverse pattern of results when habituated to an inanimate rod that repeatedly reached out and contacted one of the objects (Woodward, 1998). Again, only the actions of intentional agents are treated as goal oriented.

Consider now the case in which a child hears an adult use the word *dog* and infers that the adult intends to refer to a dog. For the child to figure out that she can also use the word in the same way (that it is a Saussurian sign), perhaps precisely the same sort of inference as above is needed:

Scene:	An adult says "dog" while looking at a dog.
Goal:	Referring to the dog
Action:	Saying "dog"
Inference:	A reference to dogs can be established by saying "dog."

Once a child believes that the adult's use of the word *dog* was used with the intent to refer to a dog, she can infer that if she herself has the same intent (to refer to a dog), then she could use the same means (saying "dog") to satisfy this goal. Just like touching a switch with your forehead turns on a light, saying the word "dog" refers to dogs.

From this perspective, it should be the asymmetrical cases that are hard for children to learn, since something has to block this inference from action to goal. Nobody has ever tried, but it should be terribly difficult to teach children a communicative system that is not Saussurian. (By the same token, it should be just as hard to teach children a noncommunicative asymmetric mapping, in which they observe someone perform an action that achieves a desirable goal but somehow infer that they themselves cannot achieve the same goal by doing the same action.) Note that the examples of asymmetric communication systems that do exist either will involve creatures who do not have a full-blown theory of mind—such as dogs and computers—or are cases that are not actually intentional. If Dad gasps when he sees a wasp, this is properly viewed *not* as a goal-directed action but as an involuntary reaction, and so children do not make the inference that this is an appropriate action to take on encountering wasps.

This theory of mind analysis makes a prediction, which is that children should start off treating *any* communicative act as a potential word. This is in contrast to the prediction that comes from the Hurford view that children's initial assumption about words is part of a special language capacity and hence that only those symbols that are linguistic units (with phonology, morphology, and syntax) should be treated this way.

Some recent studies bear on this issue. Namy and Waxman (1998) taught babies (who were learning English and not a sign language) novel gestures as object labels. For instance, the researcher would show a baby a toy apple, say "We call this," and then produce a novel gesture with her hand. In the test phase, the experimenter would then show the baby another toy apple and a different object, such as a toy pig, and ask "Can you get [the gesture]?" Woodward and Hoyne (1999) used a noisemaker that makes a squeak. The experimenter would show the baby a novel object and say something like "Look at this. [squeak]. Yeah, see it? [squeak]. Wow, look! [squeak]." In the test phase, a second experimenter would show the baby both the original object and a different object and ask "Can you get one of these? [squeak]."

Both studies found that the youngest children who were tested (18-month-olds in the gesture study and 13-month-olds in the noisemaker study) remembered the novel mapping. But the older children who were tested (26-month-olds in the gesture study; 20-month-olds in the noisemaker study) did not. A third study, by Baldwin, Bill, and Ontai (1996), used a gaze-following paradigm and found that neither 12-month-olds nor 18-month-olds treated an adult's sigh in the same way that they treated a word that the adult uttered ("dax").

These findings suggest that an understanding of words develops along two tracks. One is a notion of word that corresponds to a phonological unit, such that *dog* is a word but a squeak or gesture (for children learning a spoken language) is not. Innate language-specific expectations may be held about words in this sense that concern their phonological, morphological, and syntactic properties. Another notion of word corresponds to a Saussurian sign. This emerges from theory of mind, and hence, initially, any intentional communicative act is treated as a Saussurian sign (a phonological string, a gesture, or a sound made by a noisemaker—but not an accidental sound such as a sigh).

The results from the above studies suggest that not until some time after 18 months do these two notions come together and that children realize that only the phonological word is typically used as a Saussurian sign. They learn that people name things with phonological words, not with gestures or squeaks, and only then is their understanding of communication fully integrated into their understanding of language.

Word Learning with and without Theory of Mind

A focus on theory of mind makes some strong predictions about how certain disorders should affect the course of word learning. One central case is that of autism, a developmental disorder that affects about one in a thousand children. It is characterized by a range of deficits, including impairments of socialization, communication, and imagination. One theory is that this cluster of deficits is the product of a delayed, impaired, or nonexistent theory of mind (Baron-Cohen, Leslie & Frith, 1985). This elegantly explains the range of specific problems that autistic children have, including difficulties with understanding false belief, deception, and ignorance, while at the same time accounting for preserved abilities in other domains. (It should be noted, however, that certain other facts about autistic individuals, such as their excellent rote memories and preoccupation with parts of objects, cannot be explained in terms of this theory of mind deficit; see Happé, 1996.)

One proposal, defended in detail by Uta Frith and Francesca Happé (1994), is that the linguistic impairments of autistic individuals are not due to an additional deficit that is special to language, but exists because autistic children are impaired in the theory-of-mind abilities necessary for normal language learning.

Autistic individuals differ profoundly in their linguistic abilities. On one extreme are those with no language. About 30 percent of the individuals who are labeled autistic fall into this class. They might first appear to be deaf, since they often fail to orient to speech, and they sometimes produce odd vocalizations that do not resemble speech or babbling. On the other extreme are those who come to talk, as Asperger (1944) put it, like "little professors." They might have perfect syntax, but their prosody is bizarre (often either monotone or sing-song), and so is their pragmatics. Their language is highly literal. When asked

"Can you pass the salt?," for instance, such individuals might answer yes but do nothing. Most relevant for our purposes here, their vocabularies are said to be normal. Some of them have Asperger's syndrome, which includes only those autistic individuals who show normal language development.[1]

Falling between these two extremes are the majority of autistic individuals, who have some limited language skills. They might echo back words and phrases, either immediately or after a delay. They will show pronominal reversal, using "I" for "you" and vice-versa, and might use entire phrases in such a parroted way—for instance, saying "Do you want a biscuit?" to mean "I want a biscuit." Words and phrases are used in a "simple associative way" (Frith & Happé, 1994), so that "Apple" might always mean "Give me an apple."

Some of this odd linguistic behavior can be readily explained in terms of associative learning mechanisms. In an often cited anecdote, Kanner (1943) reported an autistic boy who used the phrase "Peter eater" to talk about saucepans. His mother explained this by recounting that when he was two years old, she was reciting the rhyme "Peter, Peter, Pumpkin Eater" to him when she dropped a saucepan with a loud clatter. Baron-Cohen, Baldwin, and Crowson (1997) discuss an autistic toddler who would call a toy truck "a sausage," apparently because his mother had said "Tommy, come and eat your sausage" as the boy was looking at his truck. These examples may indicate the hazards of learning language through a strictly associationist learning mechanism.

To explore this issue, Baron-Cohen, Baldwin, and Crowson (1997) studied autistic children using the discrepant-looking paradigm of Baldwin (1991), in which an experimenter looks at one object and utters its name while a child is attending to another object. A purely associative mechanism would lead children to map the word onto what they are attending to, while learning based on theory of mind would lead them to map the word onto what the experimenter is looking at. Autistic children made associatively based mapping errors, while both normal children and mentally handicapped children, matched to the autistic group in mental age, did not. This supports the view that these autistic children's difficulties in word learning are due to their deficit in theory of mind; they lack the inferential capacities that come naturally to normal children who are younger than two.

But so far we have explained only a subset of the autistic population, the middle group, those who have some limited success at learning language but who make unusual errors. But what about the children on the extremes? Some learn no language at all, and, more puzzling for the account here, some show surprising success at language. There are even autobiographies written by autistic individuals such as Temple Grandin. How can this be explained?

One possibility is to appeal to factors independent of theory of mind. Autism is often associated with severe mental retardation and other deficits, perhaps including specifically linguistic problems. It is plausible that these problems—in addition to a deficit in theory of mind—explain those autistic children who remain entirely mute. Harder to explain is the existence of autistic individuals who are almost normal in their language. One proposal is that a continuum of theory-of-mind abilities runs from a severely autistic individual to a normal unimpaired person. This is the approach taken by Frith and Happé (1994), who observe that those individuals who have relatively preserved language skills are the same individuals who tend to perform well on tasks designed to tap their understanding of the thoughts of other people.

In general, then, the extent of the language deficit found in autistic children may be a direct function of the severity of the theory-of-mind deficit. A severe theory-of-mind deficit might leave children without the ability to orient preferentially to speech, share attention, or follow eye gaze, and they might never be able to grasp the notion of an arbitrary sign, leading to no word learning at all. A less severe impairment might make word learning possible but limited and idiosyncratic. And in some cases, the theory-of-mind impairment might be sufficiently mild so as to leave word learning fairly unimpaired, although such individuals might still have problems with aspects of language such as irony and metaphor.

The autism research bears on an alternative perspective on the role of theory of mind in word learning, proposed by Sperber (1997) and others. Suppose that children's ability to understand the communicative intentions of other people really is essential for word learning. Still, it might be that this ability does not arise from a more general capacity to reason about mental states but instead comes from a specialized system that is part of language itself—not a theory of mind, but a *theory of communication.* This would be consistent with the view that word learning is dependent on modular systems that are special to language learning.

The weakness with this theory is that the same inferential capacities relevant to communication apply as well to noncommunicative situations. As discussed above, the same capacity required to appreciate the bidirectional nature of language (Saussure, 1916/1959) underlies the understanding of goal-directed action in general (Meltzoff, 1988). The same direction of gaze cues that children use when figuring out what object someone is labeling (Baldwin, 1991) are used to figure out which object an adult is disgusted by (Baldwin & Moses, 1994). The same ability to distinguish accidental versus purposeful action that is involved in word learning (Tomasello & Barton, 1994) extends to children's choices about which actions to imitate (Meltzoff, 1995). It is highly suggestive that these noncommunicative aspects of theory of mind emerge in normal children at roughly the same age as the communicative aspects of theory of mind, which is also the point at which children begin to learn words.

And consider autism. If a theory of communication were distinct from theory of mind, then we would expect to find autistic children with severe problems understanding the *non*communicative actions of other people but with nonetheless normal language and communication skills. But such cases do not exist.

Williams syndrome (WS) provides an interesting contrast with autism. Like autism, it is a severe disorder of genetic origin. It is rare, affecting about 1 in 20,000 to 50,000 live births, and involves severe deficits in cognitive skills, including number, problem solving, and spatial cognition. Interestingly, language (along with face processing) is relatively spared (e.g., Bellugi, Marks, Bihrle & Sabo, 1988). Despite their low IQs, the vocabulary size of individuals with WS is typically closer to their chronological age than their mental age. While developmental differences do appear in word learning between WS children and normal children (Thal, Bates & Bellugi, 1989; Stevens & Karmiloff-Smith, 1997), the outcomes are quite similar—and very different from most autistic children.

The relevant point here is that people with Williams syndrome are highly social and appear to possess a fully functioning theory of mind (Karmiloff-Smith et al., 1995). Because of their retardation, WS individuals have difficulties in domains that autistic children often succeed at, such as learning to read. But the reverse pattern shows up for language. Because of their social capacities, language is relatively spared in WS—just as it is impaired in autism. If many autistic individuals show how damaging it can be to word learning to have an impaired theory of mind, many people with WS show how a preserved theory of mind can sustain language development in the presence of other problems.

The Adult's Theory of Mind

The starting point for countless discussions of word learning is Quine's "gavagai" example: a child hears a novel word and must figure out, from an infinity of possibilities, the correct meaning. The moral that is typically drawn from this is that constraints or biases must exist. The scope, nature, and origin of such constraints is a matter of debate, but their necessity cannot be in doubt.

Katherine Nelson (1988, p. 240) suggests that this is all wrong: "The typical way children acquire words in their first language is almost completely the reverse of the Quinean paradigm. Children do not try to guess what it is that the adult intends to; rather they have certain conceptions of these aspects of the world they find interesting and, in successful cases of word acquisition it is the adult [at least in Western middle-class societies] who guesses what the child is focused on and applies an appropriate word."

This is an intriguing perspective, and it shows that alternatives do exist to viewing word learning as an inductive process. Parents could notice what a child is observing, such as a dog, and then produce the word for it, "dog." Given that children can form associations, this could be the foundation of word learning. Similarly, Lois Bloom (1993) has suggested, as part of her principle of relevance, "Words are learned when they are relevant to what the child has in mind."

There is still an emphasis on theory of mind here, but it is the *adult's*, not the child's. There is some appeal to this view. After all, one can question how good one- and two-year-olds really are at successfully inferring the thoughts of others. But nobody doubts that adults can, at least some of the time, figure out their child's thoughts and that they are motivated to help their children learn words.

Furthermore, abundant evidence exists that Western parents tailor their use of words to accord with their children's mental states. When interacting with young children, they tend to talk in the here and now, adjusting their conversational patterns to fit the current situation. They engage in "follow-in" labeling, in which they notice what their babies are looking at and name it (e.g., Collins, 1977; Golinkoff, 1986). They even seem to have an implicit understanding that children assume that new words referring to objects will be basic-level names, such as *dog* or *shoe,* and so when adults present children with words that are not basic-level names, they use linguistic cues to make it clear that the words have a different status. For instance, when adults present part names to children, they hardly ever simply point and say "Look at the ears." Instead, they typically begin by talking about the whole object ("This is a rabbit") and then introduce the part name with a possessive construction ("and these are his *ears*") (Masur, 1997; Ninio, 1980; Shipley, Kuhn & Madden, 1983). Similar linguistic support occurs for subordinates ("A *pug* is a kind of dog") and superordinates ("These are *animals.* Dogs and cats are kinds of *animals*") (Adams & Bullock, 1985; Blewitt, 1983; Callanan, 1985; Poulin-Dubois, Graham & Sippola, 1995; Shipley, Kuhn & Madden, 1983).

What role does this sort of support play in word learning? Under a strong version of the parent-centered view, children wouldn't be able to cope without it. But this is unlikely. All the experimental evidence reviewed above shows that children can learn words when the condition of preexisting joint attention is not met. Furthermore, neglected and abused children, raised in situations in which nobody is trying to teach them language, nevertheless come to know the meanings of words. And even in the happiest of families, words are not always used to refer to what children are attending to, and yet serious mapping errors are nonexistent. Finally—as Nelson herself parenthetically notes in the quotation above—these naming practices are not universal; in some societies adults make no effort to teach the meanings of words to their young children.

On the other hand, it is unlikely that all of this careful behavior on the part of adults is an utter waste of time. The argument so far has been that children are remarkably good at figuring out the thoughts of adults; this is the engine that drives the word-learning process. Is it really plausible, then, that *adults* are so inept at figuring out the thoughts of their children that they go through elaborate efforts that have no effect at all? It is more likely that parents know what they are doing and hence these strategies really do help children learn words.

This assumption has considerable support. One does not need to do a controlled study to learn that the best way to teach children a new object name is to make sure they are paying attention to the object when the word is used. Verbs are different; children actually find it harder to learn a novel verb when it is used to comment on an already ongoing event; they do better when the verb is used immediately before the event (Tomasello & Kruger, 1992)—which is exactly when parents tend to use novel verbs (Tomasello, 1992). Studies with novel categories find that children's natural assumption when exposed to a novel object label is to interpret it as referring to a basic-level object kind, which is again exactly in accord with how adults use such words (Horton & Markman, 1980; Markman & Wachtel, 1988; Mervis & Crisafi, 1982).

All of this suggests that these naming patterns on the part of adults are useful but that they just aren't necessary. The important thing to realize here is that nothing about adult naming behavior is special to dealing with children. Imagine that a college student from another country is living with you and speaks no English but wants to learn. You might slow down your rate of speech, talk about the here and now, point to objects and name them, and so on, just as you would do for a young child. If you wanted to teach the student the word *handle,* you would not pick up a cup and say "Handle"; you would more likely say something like "This is a cup, and [touching the handle] this is the handle." If you wanted to teach the word *animal,* you would not point to the dog and say, "This is an animal"; instead you would say something like "This is a dog. A dog is a kind of animal," or you would wait to find a heterogeneous group of animals and say "These are animals."

Most children raised in Western societies are in the situation of the student: they are surrounded by people who want to teach them words and who are pretty good at doing so. Children in other cultures are in the position of a student who is surrounded by people who love him, find him adorable, but have no interest in teaching him how to speak. Adults can learn words even in this more impoverished environment—and so can children.

This leads to the prediction that children raised in environments in which this support is present should learn words *faster* than those raised in other environments. If not, then most Western parents really would be wasting their time. In fact, there is a correlation between the extent to which parents engage their children in joint attention interactions and their rate of word learning (Akhtar, Dunham & Dunham, 1991; Harris, Jones & Grant, 1983; Tomasello & Todd, 1983), which is consistent with the view that parental naming behavior affects vocabulary growth, though (as discussed in chapter 2) there are other explanations for such correlations. Furthermore, anecdotal evidence shows that children raised in societies without object labeling do learn words somewhat slower than those raised in most Western societies (Lieven, 1994), though as yet no systematic research has been done into this issue.

In sum, adults' attempts to teach children words might help speed up the word-learning process. But they are not necessary for word learning and, even when they are present, do not substitute for the child's own ability to infer the referential intentions of others.

Word Learning in Chimpanzees

Another reason to believe that adult naming practices cannot entirely explain word learning has to do with species differences. This issue of precisely what nonhuman primates can and cannot learn is a controversial one, but some facts are clear. The signal systems of primates in the wild, such as vervet monkeys, are based on a small, fixed number of signs with determinate meanings; individuals do not create new words and do not teach words to their offspring (e.g., Cheney & Seyfarth, 1990). And no matter how supportive an environment they are placed in, nonhuman primates do not learn words in the same way that human children do. Even by the most enthusiastic claims, chimpanzees who are trained in the use of sign language come to use about 200 to 300 signs after years of extensive interaction (e.g., Savage-Rambaugh et al., 1993). As Lila Gleitman once remarked, if a child ever learned language the way that an ape does, the child's parents would run screaming to the nearest neurologist.

Why are chimpanzees so bad at word learning? Under a theory that sees the capacity for word learning as part of a dedicated language faculty, there is a ready answer: chimps have not evolved this language faculty, and so they do not learn words for the same reason that they do not learn phonology, morphology, and syntax. But I have suggested here that there is no special faculty for word learning, that it emerges from a host of other capacities that humans possess. So why are chimps—who are smart and capable animals, superior to one- and two-year-olds in many ways—so much worse at word learning?

One possibility is that despite their quite rich social abilities, they lack relevant aspects of theory of mind. In particular, they lack an instinctive understanding of referential intent. This is the position taken by Tomasello (1998). He notes that chimpanzees in the wild never show, offer, or point to objects for other chimpanzees. And while you can train them to point to direct their trainers to food, they never quite get the hang of it; when they see someone else point, they are mystified (Call & Tomasello, 1994; Povinelli et al., 1997). In a nice turn of phrase, Tomasello summarizes the species difference: "children use symbols, whereas other primates use signals." Under this view, a chimpanzee is in the position of the most autistic of autistic children, never understanding how words work.

Just as with babies, it is always risky making claims about what nonhuman primates cannot do; if you wait long enough, some clever researcher will prove you wrong. Moreover, chimpanzees plainly *do* have certain capacities related to theory of mind, such as the capacity to categorize certain social relationships and to conceive of conspecifics as animate and goal-directed entities. But Tomasello's hypothesis is appealing in a couple of ways.

First, it is not obvious what *other* capacity chimpanzees lack that could explain their failure to learn words. They have fine perceptual and motor skills, they are excellent at associative learning, and they seem to have the right sort of conceptual understanding of the external world. Theory of mind is the only area, other than language itself, in which they are manifestly inferior to human children.

Second, this proposal captures an insight about the behavior of nonhuman primates in the wild. As Tomasello notes, chimpanzees communicate to regulate dyadic interactions such as play and sex, but they don't communicate about other entities; they don't *refer*. This limitation in theory of mind extends to noncommunicative contexts as well, as they do not spontaneously imitate the goal-directed behavior of others. In all of these ways, a chimpanzee is profoundly different from a human without language, such as a baby or an aphasic adult. This could explain why they can never fully engage in the process of word learning.

Limitations of Theory of Mind

Could theory of mind be the whole story of word learning? Perhaps learning the meaning of a word just reduces to intentional inference; once we know how children divine the intentions of others, there is nothing left to explain.

But a lot more is needed. No matter how good they are at understanding the minds of others, children cannot learn a word without the ability to grasp the associated concept. Suppose, for the sake of argument, that two-year-olds have the same theory of mind as adults. Still, two-year-olds will not be able to learn words such as *modem* and *stockbroker* (even though these refer to observable middle-sized objects) because they don't yet know what such categories are. What theory of mind does for children is enable them to establish the mapping between a word and a concept. But this presupposes the availability of the concept.[2]

And there is still Quine's problem. Putting aside exotic possibilities such as undetached rabbit parts, how do children know whether the adult is describing the individual rabbit (as in "This is Flopsy") versus describing the rabbit as a member of a category (as in "This is a rabbit")? Pointing and eye gaze are useless here, as the overt behavior of someone using a proper name is indistinguishable from the behavior of someone using a common noun (Baldwin, 1995). So unless children are *literally* mind-readers, they have to use some other cue to tell the difference. Candidate proposals as to how this is done include a sensitivity to syntactic cues, an understanding of what sorts of things get proper names, and information gained by hearing a word used on multiple occasions (see chapter 5).

Consider also children's default expectations as to what words mean. When shown a new object and given a word for it, children are prone to think that the new word is a name for the whole object, as in *rabbit*. It is conceivable that this bias is product of theory of mind; children favor the object interpretation because they think that objects are what adults are most likely to refer to. This is a reasonable hypothesis, but in the next chapter, I present evidence suggesting that it is mistaken. For one thing, it turns out that the "object bias" is relatively impervious to pragmatic factors. It is very hard to teach a solid substance name like *wood,* for instance; regardless of what an adult says or does, young children tend to think the word is an object name. For another, the same whole object bias shows up in noncommunicative domains such as tracking and enumeration.

What is the proper place, then, of theory of mind in an account of early word learning? One way to look at it is that children use inferences about the referential intentions of others to create arrows, or pointers, from words to the world. A child hears the word "rabbit" and uses a speaker's direction of gaze to figure out what he or she is referring to. In the child's mind, an arrow is now going from *rabbit* to a rabbit. This understanding is necessary to learn the word. But the point of the arrow does not touch a concept or meaning; it touches an object in the world, a rabbit. It is up to the child to figure out from this what the word means. Does it refer to the kind (*rabbit*), a specific individual (*Flopsy*), or a property (*white*)? If the word refers to the kind, what other objects belong to the same kind—that is, what other objects are rabbits? Even with a full-blown theory of mind, the child's problems have just begun.

Notes

1. Though the criteria for "normal" here are quite liberal. For instance, Francesca Happé has pointed out to me that if a child produces only one word at the age of 24 months, this would likely count as normal development for the purposes of diagnosis. Little is known about the early language development of children who are later diagnosed with Asperger's syndrome.
2. It also presupposes the availability of the *form* of a word, which requires the ability to segment the speech stream—also a separate capacity from theory of mind.

BLANK PAGE

Chapter 4
Object Names and Other Common Nouns

The first words of a child of Harvard graduate students in the 1960s are much the same as those of a child learning French in the Paris of the 1920s or one currently learning Kaluli in the highlands of Papua New Guinea. Many early words refer to middle-sized objects—things that can move and be moved. These include names for specific people (*Mama, Dada*), animals (*dog, cat*), toys (*ball, block*), articles of clothing (*sock, shirt*), and other artifacts (*fork, chair*). There are names for substances (*juice, milk*), names for parts, typically body parts (*nose, foot*), modifiers (*hot, more*), words that refer to actions or changes (*up, allgone*), and routines that are linked to certain social interactions (*bye-bye, peek-a-boo*) (see Clark, 1993, for review). Soon afterward, verbs appear (*go, make*), as do prepositions (*in, on*) and more abstract terms (*kitchen, nap*).

This chapter addresses the question of why children's early vocabularies are the way they are. The initial focus is on object names, but it also discusses nominals in general, including those that refer to substances, actions, parts, and collections. The chapter concludes by shifting attention away from meaning and toward form, asking how children figure out which words in an utterance are the relevant ones to attend to. This problem is largely ignored in discussions of word learning, but it is a serious one, and its solution leads to some rather surprising conclusions about the sort of experience necessary for successful word learning.

Explaining the Words People Know

Learning a word involves mapping a form, such as the sound "dog," onto a meaning or concept, such as the concept of dogs. This perspective leads to three considerations underlying why children and adults know the words they do.