

## **\_Learning Language\***

Ruth Garrett Millikan

### **ABSTRACT**

Interpreting what you hear through the medium of speech sounds is in many ways just like interpreting what you see through the medium of the structured light that strikes your eyes. Understanding descriptive speech is closely akin to perception of the world through vision, ordinary hearing or touch. This, plus a flexible view of the different ways in which one mind can take account of another, helps to explain how children can learn tens of thousands of words without special instruction and without requiring a 'theory of mind.'

### **1. Introduction**

Many students of pragmatics and child language have come to believe that in order to learn a language a child must first have a 'theory of mind,' a grasp that speakers mentally represent the content they would convey when they speak. This view is reinforced by the Gricean theory of communication, according to which speakers intend their words to cause hearers to believe or to do certain things and hearers must recognize these intentions if they are to comply. The view rests on an underlying assumption that learning language involves associating words with things (objects, kinds, events, properties and so forth) or with concepts of these, these associations being acquired, one by one, by observing the usage of others. Accomplishing this task is facilitated, it is thought, by engaging in joint attention with speakers who are attending to the things they are talking about as they talk, and joint attention requires an understanding that others have minds that represent things.

Now, obviously, no one supposes that children require help of this kind in learning to associate the ordinary sounds of things with their looks, their looks with their feels and with their smells and so forth. I propose to explore this asymmetry to see just what it consists in and what it is worth.

But first, there are prior reasons to question whether a theory of mind is required for learning the meanings of words. I will mention two. We know that children understand a good deal of language well before they speak, and that they learn to speak quite well several years before they are able to pass the classic tests that psychologists have used to determine when grasp of the concept of belief is obtained, more carefully, when children grasp that people can have false beliefs. Deaf children

---

\* Some short passages of the essay have been revised from Millikan 2000 Chapter 6, Millikan 2004 Chapter 9, or Millikan 2005 Chapter 10, where more details on some of the claims made here can be found.

\*I am indebted to Natasha Millikan for pointing out a major problem with organization in an earlier draft of this essay.

Address for correspondence:

Philosophy U-54, University of Connecticut, Storrs CT, 06250

Email: ruth.Millikan@uconn.edu

don't generally pass these tests until about age 16 (Garfield et al 2001). The soundness of these tests has been questioned, but if validated, the result is clearly inconsistent with a Gricean model of language learning and understanding. Assuming that joint attention may indeed be an aid at least in early language learning, it is worth questioning whether it is best understood as requiring a theory of mind.

Much more disturbing is the fact that ordinary American school children acquire a vocabulary of 45,000 base words (not including inflections and not including probably an equal number of proper names) by 16 years of age (Bloom 2000, Chapter 1), Chomsky says (though this is surely hyperbole), 'about a word an hour from ages two to eight with lexical items typically acquired on a single exposure...' (1995, p. 15). That this is possible seems a mystery if each word must be learned by being paired with an independently grasped intension or meaning, somehow transmitted to the child by joint attention or by any other method. A large proportion of the child's words are learned by listening or reading in contexts far removed from what is being talked or read about, making it unclear how joint attention could be operating. Moreover, very often the concepts that are to go with these words are first acquired in picking up the words, rather than the concept coming first then attaching to the word. I know of no current theory of language learning that comes close to explaining these prodigious accomplishments. The idea that possessing of a theory of mind is involved might help explain how a first mole hill is built, but not a mountain.

Inquiring into the extent and significance of the difference between learning words for things and learning how these things look, how they sound, feel, taste, smell and so forth, I will begin by first reexamining the relation between understanding language and understanding the significance of the looks, feels, smells and so forth, that one encounters. I will suggest that there is a much closer resemblance between encountering words and encountering these more primitive manifestations of one's worldly surroundings than is generally acknowledged. In routine instances, I will argue, understanding descriptive language is surprisingly like direct perception of the world. During normal conversation, it is not language that is most directly perceived by the hearer, but rather the world that is perceived *through* language. The Gricean model of communication can be set aside, at least for the most ordinary sorts of linguistic communication. Learning a language is, in its essentials, just learning to perceive the world through one more sensory modality.

By the end, I will try to make plausible each of four proposals. The first (?2), mentioned just above, will require far the most discussion. The second (?3) I will touch on very lightly, details and argument having been supplied in previous work. The last two (?4 and ?5) are important details that bear directly on language learning, not merely language understanding.

- C Interpreting the meaning of what you hear through the medium of speech sounds is in relevant ways surprisingly like interpreting what you see through the medium of the structured light that strikes your eyes. Learning from speech is, in its essentials, a form of perception *of the world*, as direct as any other. This view can be turned to explain how prescriptive language is understood as well.
- C Learning words for things is much like learning what things look like, what

they feel like and so forth. It is learning one way they can sound. Acquiring the capacity to identify a thing by knowing what it looks like, feels like, sounds like, how to test for it, as well as by inference ? learning, that is, how to recognize more and more of its manifestations ? is acquiring a better and better concept of it. Being able to recognize (in context) the same word bearing the same meaning again often *constitutes* having a rudimentary concept of it, just as does knowing, say, what a thing looks like.

- C If possessing a 'theory of mind' requires understanding that others possess mental representations, then that others have communicative purposes can be recognized without having a theory of mind.
- C Indications of where the speaker's attention is focused ('joint attention'), and also of where a writer's attention is focused, can be used during the process of language interpretation and acquisition without the hearer's having to employ a theory of mind.

## **2. Understanding Descriptive Speech is Typically a lot like Directly Perceiving the World.**

### **2.1 Objects Of Direct Perception May Fall At Any Of Many Levels Of Distality**

It is traditional to assume that gathering information by being told things is a radically different sort of process from gathering information through ordinary perception. But exactly what are the differences? I will examine them in detail, for some are obviously very important. But before I do this, let me ask you, for a moment, to consider thinking of matters in this way.

There are many ways to recognize, for example, rain. There is a way that rain feels when it falls on you, and a way that it looks out the window. There is a way that it sounds falling on the rooftop, 'retetetetetet,' and a way that it sounds falling on the ground, 'shshshshsh.' And falling on English speakers, here is another way it can sound: 'Hey, it's raining!'

Most people will object, I imagine, that it is not rain you hear but a sentence. Or perhaps that it is only a sound that you hear? Then perhaps it is only a sound that you hear rather than rain on the roof? Consider: when the doorbell rings do you hear someone at the door, or just the doorbell? Or just a sound (maybe its not the doorbell)? Indeed, perhaps, you have tinnitus and it is only a sense datum that you hear? Or consider: is it a TV screen that you see when you watch the news rather than Obama? A pattern of ambient light rather than the TV screen? Best of all, perhaps all you see is a visual impression?

There is a long, but I'm sure wrong, tradition in philosophy that for each of the senses, there is but one level that reveals the real or direct objects of perception for that sense, all other levels being known only by inference. Often it is thought that this privileged level is known with certainty, with infallibility. The primary source of this doctrine, I think, is our double use of the various verbs of perception. We use these

verbs to describe our veridical perceptions, but also to describe what happens when we merely seem to perceive. On the one hand, we say that you can't 'see' an elephant unless it is there; on the other hand we say that if delirious you may 'see' pink elephants that aren't there. The trouble is that we lack for perception the distinction we draw so easily for discursive cognition, the distinction between knowledge and mere belief. Like knowledge, perception appears to be something you can't be wrong about yet, it seems, it must also be of something real. If you could be wrong ... perhaps it's not someone at the door, perhaps it's not even the doorbell, perhaps there's not even any real sound ... then that is not the level of perception.

This position is, I think, importantly mistaken. You can, if you like, hear or see at any of many different levels. What you see when you look at the television, for example, depends, first, on where you focus your eyes; it depends, second, on where you focus your mind, your attention. Nor is there reason to suppose a significant difference in directness of psychological processing when you move from one focus to another. None of these perceived objects are epistemological givens. All are discovered through processes of fallible neural construction. Nor is there any advance restriction on the level of distality of what you perceive directly. When watching TV, usually you are seeing at the level of the objects and events depicted; you see the newscaster frown or smile, you see him point to the map. Other times you might study the dots on the screen, noticing patterns in the static. The TV technician may be able to see through the dots to perceive exactly what part of the mechanism inside is ailing. The housekeeper sees the grime on the screen and reaches for a rag to clean it. Looking out a train window when the inside of the carriage is well lit, sometimes one can either watch the people reflected inside or watch the scenery outside with equal ease, neither possibility interfering with the other. Certainly I can hear someone at the door, and I can hear rain on the roof.

Why then, when someone says 'It's raining' don't we say that we hear rain, just as we do when we hear rain on the roof? The reason is not, apparently, that there is a difference in depth of psychological processing, that there is a crack that an inference must bridge, using further evidence in the one case but not the other. There is, for example, experimental evidence that what you are told goes directly into belief unless you do cognitive work to prevent this, exactly as with what you perceive through other media, such as light when directly seeing an object. Loading the cognitive systems with other tasks, say, counting backwards by threes from a thousand or pushing a button whenever a five appears on the computer screen, has the effect of facilitating belief fixation regarding whatever one reads, so presumably for whatever one hears said as well (Gilbert et al 1993). First believe, then evaluate if you get the chance, is the policy of one's cognitive systems both when one hears rain pattering on the roof and when one hears someone say 'It's raining.'

## ***2.2 What Reaches The Sense Organs Is Always Merely A Sign; The Qualities Of Ordinary Sensory Experience No More Present The 'Real Nature' Of The Perceived Than Do Words And Sentences.***

Still, one naturally has this reaction to the above, I think. Surely hearing someone say 'It's raining!' is not hearing rain! Surely that's not what rain sounds like! Worse, surely

the spoken word 'dog' doesn't sound like a dog, nor the written word look like a dog. What reaches the ear or eye when you hear or see the word 'dog' is, obviously, only a sign of a dog, not ? ???

The challenge is how to fill in the blank to articulate the intended contrast. For what reaches the ear when you hear a dog (say, its bark) is also only a *sign* of a dog, not a dog, and what reaches the eye when you see a dog is only a *sign* of a dog.<sup>1</sup> Is it that dogginess itself, doggie essence, is somehow contained in the sight of a dog or in the sound of it's bark but missing when you encounter the word 'dog'? Certain kinds of ambient energy structures, impinging on the senses, produce experiences that transparently reveal dog nature, dog-in-itself-ness, whereas others do not?

That can't be right. The subjective 'feel' of the information arriving at one's senses, in its many varieties of vehicles, the various textures caused by the proximal stimulations carrying information on the whereabouts of a dog, are not relevant. Perhaps subjective shape appearances, subjective shape qualities, are different for bats and dolphins who *hear* shapes than for us, but surely this wouldn't prevent them from directly perceiving shapes. Nor does the feel of a shape reveal its essence more or less transparently than it's look. (Which one gets it right?)

What does seem to be true, and important, of much ordinary perception, however, and to be especially obvious in the case of *seeing* the dog, is that certain variations in structure of the light impinging on the retina, certain kinds of transformations of that structure, would correspond pretty systematically to certain kinds of variations of that dog, or of whatever distal object was the light's source. The structure of the light would vary, not just with a dog being there or not being there, but according to variations in a good number of different properties of the perceived object. The word 'dog' has no such significant structure. We have to focus on variations in whole phrases and sentences to understand how language carries information ? rearrangements of words or substitutions of words for other words. Hearing the word 'dog' in isolation definitely is not hearing a dog! But hearing 'There's a dog scratching up your lawn' might be hearing a dog scratching up your lawn. That's the idea I want to explore.

First I'll discuss two important differences between ordinary perception, and acquiring information through language. I'll argue that though important, neither implies a difference in directness of psychological processing. Next I'll compare the information received through ordinary perception and through language both with regard to its abstractness and with regard to the 'filling in' required for interpretation. Then I'll be ready to suggest implications for language learning.

### ***2.3 Normally, The Spatial And Temporal Relation Of The Perceiver To The Object Perceived Is Given In Ordinary Perception, But Not When Believing What One Hears Said***

If you see a dog, you normally also see its spatial relation to you, and whatever you perceive it as doing is done at the time you perceive it. But if you hear someone talking

---

<sup>1</sup> The bark is a natural sign of the dog whereas the word 'dog' is a conventional sign. In both cases, granted the sentence containing 'dog' is true not accidentally but for normal reasons, the sign is a manifestation of a dog. In my (2004, Chapters 5 and 6) I examine the relation between 'locally recurrent natural signs' and linguistic signs, showing that linguistic signs that are true or satisfied in accordance with a normal (technically, 'Normal') explanation are a kind of locally recurrent natural sign.

about the dog, typically you do not hear about the dog's spatial relation to you, nor about the temporal relation the dog-doings told about have to you. This marks a very important difference *in the content perceived* in the two circumstances, but not a difference in directness of psychological processing. Compare also some intermediate cases that fall between ordinary perception and gathering information through language. When you watch television, the spatial relation to yourself of what you are perceiving is not given nor, unless the program is live, is the temporal relation. Yet you perceive that Obama frowns or smiles just as immediately as you would in his presence.

The important difference in content may, however, help to explain the difference expressed in the syntax between, say, 'hearing that the dog is barking' when someone has said 'the dog is barking,' and 'hearing the dog bark.' Before the advent of modern technology (which was when the language grew up) you either perceived something close, at the same time perceiving its relation to you, and often with a great richness of detail (and redundancy...see ?2.5 below), or you perceived it through language, often when it was not close, without perceiving its relation to you, and without such great detail (or redundancy). Only the second involved 'perceiving,' for example, 'hearing,' the first merely a 'hearing that.' Nowadays, sensing that this sharp difference has been blurred, we would usually say not just 'I saw Obama' (nor merely 'I saw that Obama....') but rather 'I saw Obama *on TV*.' (We might, however, say 'I heard Obama's speech,' omitting '...on TV.')

The important point to be grasped here is that changes from the least mediated kinds of perception to more mediated kinds, involving more stages, more complexity in the media, greater causal distance as it were ? for example, to the gathering of information through linguistic signs ? are, as such, changes in the length and complexity of external links to the perceived, not of internal links. It is a mistake simply to assume that degree of complexity in external information links necessarily corresponds to degree of complexity in internal processing.

#### **2.4 Ordinary Perception Is Much More Reliable Than Testimony**

The second feature that clearly distinguishes ordinary perception from learning through language is its reliability. You can, of course, be wrong that you hear someone at the door. But it is much more difficult to deceive people about what they are directly seeing or hearing close by, whereas it is easy to deceive with language. Given a modern understanding of the mechanisms of perception, however, and a substantial technology, it is possible materially to fool the human ear and eye. False appearances can be arranged in the laboratory. And false appearances are now easily arranged using modern communications media. Though generally overlooked in this connection, the media offer much the most common current illustration of the persistence of perceptual illusion. (After seeing her daddy on television, the small daughter of a friend asked him, 'Daddy, how did you get in there?')

Persistent perceptual illusions are also very easily arranged through language. Sentences that are false continue to present the same false appearances even when you know they are false. They do not then shift and appear to mean something different. Compare: in water, oars look bent and the reflections of the trees show them moving in ripples, even though we are not at all tempted to believe that our oars are actually bent or that the trees are actually moving in ripples. Exactly similarly, a sentence keeps

saying the same thing to you, even when you know it is false. Coming to know the world through language is in this way continuous with coming to know it through ordinary perception. Although gathering information through language is much more fallible than gathering information through ordinary perception, again, the difference is one of degree not of kind.

### **2.5 Both Ordinary Perception And Language Present Only Abstract Information**

Besides the two real and important differences just mentioned between ordinary perceptual belief and believing what you hear through language, it is commonly alleged that the objects of perception are always perfectly concrete, fully filled in, whereas the information given in language is irremediably abstract and sketchy. Much of modern pragmatics rests on the assumption that a hearer must fill out the bare sketch presented to him in language to yield the richer understandings typically communicated in conversation, and that this is done using inference about the probable intentions of speakers. The hearer must make inferences about what is in the speaker's mind, something obviously not required during ordinary perception of the world. But I think this is a false contrast. First, ordinary perception itself is highly abstract. Second (?2.6 below), just as in the case of language understanding, much that is ordinarily perceived is not given in the actual data that informs perception but is filled in during perceptual processing. (Obviously this 'filling in' does not involve considering anyone's intentions.) Similarly, I suggest, for the case of perception of the world through language.

The abstractness of ordinary perception is illustrated by occasions on which one can easily perceive the presence of a determinable property without perceiving its determinate value. For example, you can see that an object is some shade or other of red but would need a better light to tell which shade. Or you may clearly perceive that some large object is moving over there in the dark but not be able to perceive what is moving. Striking testimonies to the abstract nature of perception are certain illusions of motion, for example, the waterfall illusion (Crane 1989) which presents to the perceiver a landscape that is at once clearly moving and clearly staying in exactly the same place. It moves rapidly, but not from one place to another. (Contradictions, presumably, are found only in abstractions.) That the world is presented through language very abstractly, does not then argue against the essentially perceptual nature of the presenting.

Gunnar Björnsson has pointed out to me that one thing making seeing a thing seem so very different from hearing about it is the enormous richness of detail and also redundancy of information you receive through vision, as you move eyes, head and body about to gain different perspectives. (J.J. Gibson pointed to this as an essential difference between perception and, say, seeing in a photograph.) Vision thus seems to have a dead certainty about it. In touch, also, you explore and pull in redundant information. These are indeed important differences from learning through language, but differences of degree not of kind.

### **2.6 Both Ordinary Perception And Language Understanding Involve A Great Deal Of Interpretation And 'Filling In' By The Perceptual Systems**

T/AE C/AT IS ON T/AE M/AT

Do you directly see the /A as an 'a' and only infer it as an 'h,' or is it the other way

around? Or do you have to make inferences in both cases?

Ordinary perception is not only abstract, it is interpreted and filled in, apparently by pre-inferential processing, way beyond what is dictated by currently incoming data. It is filled in using context (maybe think connectionist nets). An animal that is nearly completely occluded behind vegetation may none the less be identified easily if its color contrasts with that of the foreground or if it is moving. Similarly, half a sentence or even a single word, in the right context, often can convey a full piece of information. For a classic example, when someone calls 'I'm ready,' one generally knows for what they are ready.

If a surrogate dime that is distinctly larger than a real dime is placed at a distance in front of you, in the absence of clear depth cues you will perceive the dime as dime-sized but as further away than it actually is. That is, both size and depth are perceptually filled in according to context rather than 'actually seen.' If you listen to the syllable 'ga' while watching the apparent speaker saying 'ba' you will compellingly hear the syllable 'da' (McGurk and MacDonald 1976). If tiny lights are attached to a person's main joints and a video made showing only these points of light as the person walks about in the dark, watching the video one immediately and distinctly sees a person walking about (Johansson 1975). A circle with two marks above inside and a line below is seen as a face, the angles of the marks and curvature of the line determining whether the face is happy, angry or sad. Stick figures are readily seen as people or as animals, standing, walking, running, lifting things, boxing, waving good-bye. If you see one of Wittgenstein's duck-rabbits surrounded by drawings of ducks you will see it as a duck, but if surrounded by rabbits then as a rabbit. If you draw water in surrounding it, you will again see it as a duck. Consider how you may immediately see a certain squiggly line in a cartoon strip to be shoes in the character's hand rather than, say, a bottle, because the fellow is obviously tip-toeing up the stairs in stocking feet, the clock on the wall saying 2:30. None of this filling in in perception involves thinking of anyone's intentions.

Similarly, what is conveyed by language can be much more determinate than the semantic meaning carried merely by linguistic convention. When one hears that France is hexagonal, that a building is hexagonal and that a machine nut is hexagonal one interprets these hexagonals to have different degrees of perfection. The semantics of a possessive allows it to refer to any pairing relation coupling one kind of thing with another, such as ownership, physical possession, current responsibility for and so forth. 'John's book' may be the one he owns or carries, the one he wrote or bought or brought, and so forth, but, in context, you perceive immediately what kind of relation is involved. Words that have multiple senses are effortlessly disambiguated in linguistic context, as the ~~A~~ above is effortlessly disambiguated. Other forms are disambiguated just as easily given the immediate external context. One only has to know that we are playing blackjack ('Hit me!'), or that we are eating together ('Could you pass me the sugar?' is not functionally a question). To suppose that this must typically involve first having thoughts of speaker intentions is as unwarranted as for the case of interpreting what stick figures are doing, or filling out the blurrily seen animal that just scrambled up the tree as a squirrel rather than a rabbit (obviously).<sup>2</sup>

---

<sup>2</sup> An interesting example of the similarity of perception through language to



## **2.7 We Generally Fail To Perceive Language As Such**

Phonemes are not sounds, not acoustic phenomena. What is the same acoustically can sound as different phonemes, depending on context. According to the motor theory of speech perception, what you hear as the same phoneme again is the same movement or same posture aimed at by the vocal tract (Liberman & Mattingly 1985, 1989; Mattingly & Studdart-Kennedy 1991). When you identify a phoneme as a p-sound or an aa-sound you are not in fact identifying a sound at all but rather a vocal tract position. My further suggestion is that when you are listening to speech in order to gain information, you hear not just through the acoustics to the vocal tract gestures, but through the vocal tract gestures to the world. Of course, educated adults *can* attend to phonemes instead of the world if they try, or to phonetics (say, in order to place a foreign accent) or to words, to morphemes., to sentences or to grammatical forms. That is like studying the dots on the TV screen. They may also attend to the speaker's intentions. (They also may sometimes attend to the producer's intentions when watching TV.

That words and sentences are not what we usually hear is suggested, for example, by the ease with which people can remember the content of what was said ten seconds ago in contrast to the difficulty they have in remembering the words. Phoneme detection and word detection do have to take place within the cognitive systems during the process of language understanding, of course, just as light-gradient detection and edge detection take place during the process of seeing an object. But there is no obvious reason to suppose that concepts of phonemes and words are involved in the one case any more than concepts of light gradients and edges are involved in the other.

---

ordinary perception is the understanding of metaphor. Good cartoon caricatures of people illustrate that gross distortions of the right kind are sometimes more readily recognized in perception than the same items presented more accurately.

Understanding metaphor seems to be similar. It is understood effortlessly in context, usually not even noticed, though sometimes striking a somewhat humorous note just as pictorial caricatures can.

Little children usually cannot segment words into their component phonemes, or recognize phonemes as recurring entities at all, until they are five or six (Lieberman et al 1974). They also have difficulty with the very concept of a word.<sup>3</sup> Indeed, I think there is evidence that small children never perceive language at the level of words and sentences for conceptual purposes at all. Consider, for example, the following dialogue. (It has often been used to illustrate merely that small children are immune to attempts to correct their speech, hence that they don't learn to talk correctly through the imposition of social sanctions. But I think it illustrates something deeper.)

Child: Nobody don't like me

Mother: No, say 'nobody likes me'

Child: Nobody don't like me

(eight repetitions of this dialogue)

Mother: No, now listen carefully; say '*nobody likes me*'

Child: Oh! Nobody don't likes me (MacNeill 1966)

What seems obvious here (besides that the mother seems rather callous) is not that the child is resistant to correction. The child seems to be trying very hard. Instead, it seems that the child is simply is not hearing the words, any more than you usually see the dots on the TV screen. Both of you are bypassing the vehicle, focusing your powers of perception directly on the content. Compatibly, three to four year olds are typically unable to say whether they have just observed something for themselves or whether they have just been told about it, mostly saying that they have observed what they really have only been told about (Whitcombe and Robinson 2000; Gopnik and Graf 1988; O'Neill and Gopnik 1991; Wimmer, Hogrefe and Perner 1988). For the child, it seems, perception through language is just perception, observing for oneself.<sup>4</sup> Compare observing for oneself through a telescope.

## **2.8 Recognizing What's Being Talked About Involves Tracking, As Does Recognizing What You Hear Or See More Directly**

A difficulty that may arise in identifying either what is perceived or what is being talked about is that different things can manifest themselves in like ways. Words for different things may sound alike just as different objects may look alike, Objects and properties cannot be tracked through the medium of words merely by recognizing phonological similarities.

There are many people named 'Paul' in the world. Which one is referred to when someone says 'Paul' often seems to depend merely on which one the speaker happens to have in mind. Similarly, the content of quantifiers and definite descriptions is frequently determined as the domain the speaker is mentally focusing on, not just by the context of speaking. Does it follow in these cases that the hearer has to think about the speaker's intentions in order to interpret? Or are there ways of interpreting the

---

<sup>3</sup> For example (Susan Carey, private correspondence) they think 'witch' is not a word because there are no witches. (A word is a manifestation of a thing: no thing, no manifestation, not a real word.)

<sup>4</sup> On the other hand, Robyn Carston writes that there is some evidence that young children do treat what they have been told as being less reliable than what they observe for themselves (private correspondence). As Carston notes, however, this is not inconsistent with a failure to perceive linguistic forms as such. It may be that children, or that we all, also favor what has been seen over what has been felt or perceived through ordinary sounds.

speaker's words without understanding the domain of their focus as being represented in their minds? Can we read the necessity of 'understanding *what the speaker has in mind*' transparently here rather than opaquely? Employing a full-fledged representational theory of mind in order to interpret what another person is saying or meaning is always a possibility for an adult, but the question is how often this method is required.

As in the case of ordinary perception, words and sentences occur in contexts that support additional methods of referential tracking. Tracking through language can be analogous to perceptual tracking in relying on expected spatial, temporal and causal relations (compare expected trajectories) rather than on stimulus constancies. I recognize that it is *John's* elbow poking out over there behind the lamp because I saw John head that way with a book just a moment ago. I know it is Rover who is barking like that because I'm at home and Rover, who barks like that, lives next door. I assume that Johnny's rash is the measles because he was playing with Billy and Jane recently, just before they got the measles. Had he been playing with Tom and Sally who got scarlet fever I would interpret his rash differently. Similarly, if my husband announces as he returns from school that 'Paul' has sprained his ankle I will interpret him as referring to one person, to a different person if he announces it as he returns from playing tennis, to still a third person if he has come from playing music with friends. Other times I may know which Paul by which one is close at hand, by where we happen to be or what we are doing, or by knowing what my husband's past experience has been. Often I will know by the general subject of conversation. As he talks I am thinking about the corner of the world he is talking about, and I recognize things in that corner just as I recognize things in my own corner, for example, the bark of the dog next door. Thinking about the same corner my husband is thinking about is not thinking about his thinking.

### ***2.9 We Can Turn This View To Explain, As Well, How Prescriptive Language Is Understood***

Recognizing a linguistic reference to a thing, I have been suggesting, is just another way of recognizing the thing itself, recognizing it through one more medium of manifestation. Think of this medium as like an instrument that aids perception. Like a camera, a radio, a cat scan, or a microscope, another person who talks to me picks up information-bearing patterns from the environment, focuses them, translates them into a new medium and beams them at me. Or think of living in a language community as like being inundated in one more sea of ambient energy. Like the surrounding light, or like a TV set, surrounding people transmit the structure of the environment to me in ways that I have been tuned to interpret.

Yet language has other uses, of course, than description, and it is highly implausible that the way language is processed in the case of descriptives is completely different from the case, say, of directives. Taking orders is not perceiving snatches of the world. This observation prompts examining more closely exactly what is at stake in the suggestion that understanding descriptive language is much like another mode of perception. The proposal concerns the way language is processed. It concerns, paradigmatically, how the transition from auditory stimuli to understanding is effected.

The suggestion is that language stimuli are not, in the main ways that matter, significantly different from ordinary perceptual inputs. They are similarly abstract, context bound and ? this is what is least intuitive ? no less revealing of the real natures of what they manifest (being more exact now) when the language is descriptive, true, and true not by accident but for normal reasons (footnote 1 above). Now one can use demonstrations, or pictures, to indicate what is to be done as well as to indicate what has been done and there is no reason to suppose that the processing from stimuli to understanding differs in the two cases. One can visualize what one is to do as well as what one has done; presumably there is a similarity in mental representation between these. It is reasonable that there is, in general, a similarity in inner representation among understanding that one intends to do A, understanding that one will do A and understanding that one has done A. Indeed, I have argued (1996, 2005 Chapter 9) that intentions are probably 'pushmi-pullyu representations,' the numerically same representations being used in one's mental economy both as representations of what will be the case in future and as blueprints guiding one toward that future. It seems not peculiar then to suggest that the processing from linguistic stimuli to the understanding of directive or prescriptive language parallels the processing when understanding descriptive language.

### ***3. The Ability Merely To Grasp Sameness Of Referential Meaning In Different Tokens Of The Same Word Can Constitute A Budding Concept***

I want to add a bass line to this composition before moving on. In previous work I have argued for a theory of empirical concepts that can be woven in here so as greatly to strengthen and support both the proposal on language understanding above and the proposals on language learning below. The theory is a maverick one, it can only be hinted at here, nor can I do much to defend it in this short compass, but it is articulated in considerable detail in my (1984 Chapter 15, 1996, 1998, 2000, 2009, 2010).

Unless you can recognize the same thing again when you encounter different manifestations of it, you can gain no inductive knowledge of it, nor apply any knowledge you already have of it, nor can you acquire procedural skills in using it. But the proximal simulations that actually impinge on your sensory surfaces carrying (locally recurrent) natural information (Millikan 2004, chapters 3 and 4) about the same thing are, to a first approximation, never the same twice. The view I espouse is that what an empirical concept of any X IS, is the ability to reidentify X ? something objective in the world such as an individual, a property, a real kind or an empirical relation ? by recognizing manifestations of it in the structured energies that impinge on one's senses. Commonly, one will learn to recognize X as manifested in a great variety of ways, very distally and quite proximally, through many and diverse media that affect various of one's senses, by probing and exploring, and also by inference. Conceptual abilities of this kind are used for collecting together information about X that has been gathered in many different ways making it possible to acquire procedural knowledge, to make inductive and deductive inferences and to apply one's knowledge. Summarily, a concept of X enables gathering natural information about X that arrives at the senses in a variety of

stimulatory forms into one focus, so as to bear together on theoretical or practical learning and knowledge use.

The most important detail of this proposed theory of empirical concepts for our purposes here is this. Recall that the sight of a dog ? rather, any of the myriad different kinds of sights indicating a dog ? does not carry with it any special insight into doggie essence, nor does the sound of a dog ? rather, any of the myriad different kinds of sounds indicating a dog ? carry such insight. Further, none of the ways that one is able to identify dogs any more defines the extension of one's concept of a dog than any other. (In Millikan 2000, 2010 I explain how the natural world itself completes the determination of extension for basic empirical terms). It is an empirical matter that one's various ways of identifying dogs tend to focus on one and the same real kind, that one's various methods of identifying one's spouse or roommate tend to converge on the same individual, that one's various ways of determining the shape or the temperature of a thing tend to converge on the same property, and so forth. (The ultimate test for this is the enabling of practical learning and/or the avoidance of contradiction in thought (Millikan 2000, 2010).) Given this, there is no reason to partition off the ability to recognize through words, when these are found in the context of genuine natural information-carrying utterances (footnote 1), as intrinsically more peripheral to the concept of a thing than any other way of recognizing it. I can know something only by sight or only by certain special kinds of sights of it, or only by smell, or only by recognizing the word for it in context. Any of these may enable me to pick up more and more information about it and thus constitute a budding concept of it, in the last case, relying on the recognitions skills of others in my language community.

Recall Putnam's theory of the 'Division of Linguistic Labor' and Burge's claim that the very content of one's thought sometimes passes through the word usages of a surrounding language community (Putnam 1975a, 1975b; Burge 1986, 1988). I am offering a different explanation of the phenomena they wished to explain. Having a concept grounded only through language is no different than having a concept grounded only through any other mode of perception. Such a concept is in no way secondary. True, others must help me to have such a concept, just as a television (or perhaps just the electric lighting) may have to help me if I am to see Obama. But just as I really do see Obama on television, indeed, just as I might see him in the flesh, having a concept through language is really having a concept. It is really thinking of something.

To be able to recognize a word is to have a handle on tracking its object, tracking this object by means of its manifestations in a particular language community. For the person who remembers faces easily, one look at a new person may be enough to enable them to recognize that person again, thus to accumulate information about them over time. Similarly, for one who remembers words easily, one hearing of a new word may be enough to enable them to recognize, in context, that word with that meaning again. Thus they can accumulate information about its referent or extension over time; they have acquired a concept of this referent or extension. In this way, simply grasping the phonological structure of a language and the rudiments of how to parse it comes close to enabling one to help oneself to an embryo concept corresponding to every extensional term in the language. It enables one conceptually to track these things and pick up more information about them, which is what constitutes having an empirical

concept.

That, I propose, is the center key to why it is possible for small children to learn, as Chomsky exaggeratedly put it, 'a word an hour,' and to how can they can acquire many words along with their concepts just by listening or reading. Now for some details of that process.

#### **4. A Child Can Recognize Communicative Purposes Without Having A Theory Of Mind**

Certainly I don't think children can learn language without grasping, in some way, that other people are purposefully trying to inform them and direct them. But grasping this needn't require possessing a theory of mental representation. Purposes can be understood without understanding them as *represented* purposes. Indeed, purposes can take many forms other than the form of explicit intentions ? the purpose of the protective eye-blink reflex, the purposes of conditioned behaviors, the purposes involved in following immediate affordances (ducking the low door, reaching for your glass as you talk), the purposes usually involved (one would have thought, at least) in uncontrived speaking (to cause beliefs). Recognizing a purpose is not, typically, attributing a mental representation. Compatibly, many animals are clearly capable of recognizing one another's purposes. Dogs and cats, for example, could neither play together nor fight with each other without having some grasp of where each other's behaviors were directed, so as to anticipate each other's moves. Clearly, they often have excellent understanding of where the other is headed, what the other is aiming for.

A simple form that this understanding may take is understanding the other's behavior as 'goal directed,' that is, as part of a flexible pattern with a strong tendency to result in a given effect regardless of interference. The animal that is goal directed behaves so as to reach a certain result one way or another despite potentially deflecting forces. This has to be the understanding both of the dog and of the squirrel it is chasing. Each has to understand the other as being attracted toward certain goals, as being strongly disposed to continue a particular spurt of behavior until a certain kind of result is reached. This must also be the way a dog understands its master's intentions. Goals appear in this light as though they were future states that are actively drawing present events towards them. Aristotle spoke of 'final causes,' a term that could express this idea quite nicely.

A child can understand purposive behavior as goal directed exactly as an intelligent animal can. Studies show, for example, that a small child can identify the candy on which another child's eyes are focused and will volunteer it as the one the other child wants. The child understands toward which candy bar the other child's behavior will be directed. Similarly, the child lacking a representational theory of mind need have no problem understanding that someone else is trying to help him. When people cooperate they are directed toward the same goals and understand one another

to be so directed. The young child often finds that it's mother is directed towards the child's own goals and easily learns ways to bring this happy coincidence about. Similarly, no representational theory of mind is needed to understand that another's speech is purposefully produced and directed to some effect, and to attend what that effect might be. What part of the world is mother purposefully trying to reveal to me now, in pointing or speaking? What will the verbal picture that Mother is now drawing show? When others talk to her, the child expects to learn something about the world or, in the case of directives, to learn what to do to accomplish her goals (perhaps how to stay out of trouble). She doesn't expect to learn something about another's mind.

Speaking more generally, it is obvious, of course, that for a hearer to understand what a speaker means is for her to grasp what it is the speaker's purpose to convey. The question at issue is how frequently it is necessary that the hearer grasp 'what the speaker intends' where that phrase needs to be read opaquely, and how frequently as read merely transparently. If the hearer grasps the message the speaker is trying to impart, that is, if she grasps that such and such is the case or that she is to do such and such, that will usually be enough. The question is whether, in the process of deriving this, the hearer must usually represent to herself that the speaker has imparting this as a represented intention. In the usual case, one might suppose, if the hearer needs to diagnose the hearer's intentions she must first interpret the speaker's words (in context), not the other way around. Indeed, it is sensible to ask whether the speaker even harbors a represented intention in the usual case.<sup>5</sup>

### ***5. Indications Of The Focus Of A Speaker's Attention Can Be Used To Interpret Reference Without A Theory Of Mind***

If the proposal in ?2 is right, then for the young child, language serves roughly as another medium of perception, a medium through which to perceive the world. Exactly as the child perceives the world through her eyes without knowing anything about light, and through her sense of touch and smell without knowing anything about physical forces or chemicals, the child perceives the world through language without knowing anything about words and sentences. The child's brain (like that of a chimpanzee) is equipped at birth with a neuronal organization that becomes tuned to interpret the visual arrays around it. Similarly, the child's brain (*unlike* that of a chimpanzee) is equipped with a neuronal organization that becomes tuned to interpret the kinds of informational patterns that language presents.

As the infant matures she learns what things feel like, what those same things

---

<sup>5</sup> It may be worth noting that Grice's original form of argument (1957), designed to show that speakers must intend to cause hearers to adopt certain mental attitudes and that hearers must understand these speaker intentions, was fallacious. Grice's argument was, in sketch, that if the hearer H were to think that the speaker S did not intend H to believe what S said, then H would not believe what S said, so in the normal case it must be that H thinks that the speaker does intend him to believe. A parallel argument would be that if I thought Ben Laden had placed a land mine under my living room rug I would not walk across my living room so calmly, so in walking across so calmly, I must believe that Ben Laden has not placed a land mine under the rug. Suppose now that we apply this argument also to small children in my living room who have never heard either of Ben Laden or of land mines!

look like, what they sound like, often what they smell and taste like. She understands speakers as purposefully *showing* her things, exhibiting things to her in their speaking, and she tries to interpret what she is being shown. ?But Mamma's words are over here while the dog she talks about is way over there, or yesterday. How then does the child come to fill out what her words are about?

The child can discover what non-speech sounds mean, in large part, because hearing is directional. She has learned or innately knows to look or feel in the place or direction of the sound. In this way she soon learns to perceive many ordinary sounds as bearing information about ordinary happenings, such as doors closing, drawers opening, spoons scraping, the eggbeater beating, the toaster popping up and her own footsteps. (Notice that we could have put it the other way around just as well. She learns to understand what it is she is seeing by its direction and timing relative to what she hears and feels. Though we often think only of seeing as *really* understanding what's there, the blind do understand what's in the world too.) Similarly, the locations or directions of things felt, hefted, tasted, smelled and so forth, can often be discerned simultaneously with the direction of their sight or sound. But what's funny about language, I have said (?2.3) is that it does not generally show your own relation to the things you perceive through it. How then can the child tell which aspects of the world around her are being manifested through which speech sounds? How does she come to identify what she perceives through language with what she perceives through her eyes, touch and so forth?

When things go in the way that accounts for the survival of languages, another person who tells you things is like a pair of binoculars. They pick up patterns in the ambient energy surrounding them, interpret these patterns as manifesting certain configurations of distal objects, translate this interpreted structure into new patterns of structured energy, send them to you. You know what you are seeing through binoculars in part because you know in which direction the binoculars are pointing. Similarly, whether this is learned or innate, children grasp very early that what people talking to them are talking about is often where these people's purposive activities are directed hence, often, where their eyes or hands are directed as well. Children do not have to think about the construction of the insides of people's minds in order to grasp this, any more than you have to think about how the binoculars are constructed inside in order to tell what you are seeing through them. Thus the phenomenon of joint looking and joint attention, so much discussed in the literature.

But children learn scores of words by listening to people talk about things they are neither observing nor currently aiming toward.

Consider how you recognize what's in a photograph. Some things in the photograph you may recognize straight off, perhaps certain people, or a familiar configuration of buildings. Similarly, some of the words referring to things not currently present will already be familiar to the child. Beginning with this seed of understanding, more may be added in a number of ways. The photograph shows a real configuration of various objects and properties on which a camera was once focused. The scene was once real, so recognizing what is in one part of the picture may give context for recognizing what's in another. That's Aunt Sally in the foreground so it must be her new house in the background and the baby pulling the cap down over its eyes must be Billy.



Similarly, when part of what is being talked about is understood, then what past event or what general kind of event the speaker is focusing on may become clear, bringing that event into focus for the listening child as well. In this way, the referents/meanings of various new words may be identified with things the child can already identify through ordinary perception. Similarly, if you are talking about an event that I recognize, my mind will be focused where your mind is focused, and I will understand the proper names you use, your descriptions and the domains of your quantifiers accordingly, without any concern for what's inside your mind.

One may also know where the camera that took a certain photograph has been shooting. Since the picture of the elderly couple falls on the roll between the picture of Aunt Sally and the one of baby Billy in the bath, the couple must be her neighbors who came over for tea. Similarly, people talk about what they have observed in the past, most often about what they have quite recently observed. What people talk about typically depends on where they have been, what they have heard about or experienced. Thus the expression, when we have had difficulty understanding what someone is talking about, 'I had no idea where he was coming from!' Just as a child can understand, say, that what a digital camera shows on its monitor will be something to which it has recently been exposed, children can understand that people are often talking about what they have recently experienced. Say, Mother is talking to Daddy about what they, the child and Mother, just did in the park. Just as a child will notice what the person talking to her is looking at or what they are doing or aiming at, the child may know where this person has just been, or what they have just been talking about, or may already know about an occurrence being described. Children can do this without wielding a representational theory of mind, just as they don't need a theory of how the digital camera work inside.<sup>6</sup>

## **6. Conclusion**

The child learns to understand what she is seeing by finding out how it sounds, how it feels, tastes and smells and how it sounds through the local language, exactly as she learns to understand what she hears or hears about, feels or tastes, by finding out how it looks. No sensory modality exhibits the true nature of a sensed object more than does any other. Objects that a child has only perceived through language, objects that she has heard talked about but has never encountered through other perceptual modes, may be thought of and judged about as easily and directly as objects she has seen but not heard named. Knowing the meaning of a word in linguistic context is exactly like knowing the meaning of a sight or an ordinary sound in context. The word allows the child to re-identify the object she is receiving information about so as to put what goes

---

<sup>6</sup> Following the mental focus of another person can be like following the focus of binoculars or of a camera. On the other hand, so as to be understood reliably, speakers do also learn to leave trail markers as they shift their focus from one domain to another. They accommodate their discourse to the hearer, taking account of what the hearer might be expected already to know or to have experienced. Whether leaving trails tailored to accommodate the individual hearer requires the speaker to employ a representational theory of mind might be a good question for empirical study. For example, how good are small children at tailoring their conversation to the needs of the particular hearer, and if they are good at it, how do they do it? It would be very interesting to find out.

together into the same folder exactly as would its sight or its feel. Once initiated into the realm of language, language begins to speak to the child directly, just as sight, sound and touch do. We are not surprised that the child can learn what many thousands of things look like in just a few years. Similarly for what many thousands of things sound like through the local language.

## References

- Bloom, P. 2000: *How Children Learn the Meanings of Words*. Cambridge MA: The MIT Press.
- Burge, T. 1986: Individualism and psychology. *The Philosophical Review*, 95: 3-45.
- Burge, T. 1988: Individualism and self knowledge. *The Journal of Philosophy*, 85, 649-663.
- Crane, T. 1989: The waterfall illusion. *Analysis*, 483, 142-147.
- Chomsky, N. 1995: Language and nature. *Mind*, 104, 1?61.
- Gilbert, D. T Tafarodi, R. W., & Malone, P. S.: 1993 You can?t not believe everything you read. *Journal of Personality and Social Psychology*, 65, 221-233.
- Garfield, J., C. Peterson and T. Perry 2001: Social cognition, language acquisition and the development of the theory of mind. *Mind and Language* ,16.5, 494-541
- Gopnik, A. and P. Graf 1988: Knowing how you know; Young children?s ability to remember and identify the sources of their beliefs. *Child Development*, 56, 1366-1371.
- Grice, H.P. 1957: Meaning. *The Philosophical Review*, 66, 377-88.
- Grice, H.P. 1975. Logic and conversation. In P. Cole and J. Morgan (eds.), *Syntax and Semantics*, vol 3. New York: Academic Press.
- Johansson, G. 1975: Visual motion perception. *Scientific American*, 232, 76-89.
- Lieberman I.Y., D. Shankweiler, F.W. Fisher and B. Carter 1974: Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18:201-12
- Lieberman, A.M. and I.G. Mattingly 1985: The motor theory of speech perception revised. *Cognition*, 21, 1-36
- Lieberman, A.M. 1999: The reading researcher and the reading teacher need the right theory of speech. *Scientific Studies of Reading* ,3, 95-111.
- McGurk, H. and T. MacDonald 1976: Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Mattingly, I.G. and M. Studdart-Kennedy eds. 1991: *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NY: Erlbaum.
- Millikan, R.G. 1984: *Language, Thought and Other Biological Categories*. Cambridge MA: The MIT Press.
- Millikan, R.G. 1996: Pushmi-pullyu representations. In J. Tomberlin (ed.), *Philosophical Perspectives* vol. IX. Atascadero CA: Ridgeview Publishing, 185-200. Reprinted in L. May and M. Friedman (eds.), *Mind and Morals*. MIT Press 1996, 145-161. Reprinted in S. and G. Seay (eds.), *Philosophy of Language, The Central Topics*. Rowman & Littlefield 2008, 363-74.
- Millikan, R.G. 1998: A common structure for concepts of individuals, stuffs, and basic kinds: More mama, more milk and more mouse. *Behavioral and Brain Sciences*,22.1, 55-65. Reprinted in E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge MA, MIT Press 1999, 525-547.
- Millikan, R.G. 2000: *On Clear and Confused Ideas*. Cambridge UK: Cambridge University Press.
- Millikan, R.G. 2004: *Varieties of Meaning*. Cambridge MA: The MIT Press.
- Millikan, R.G. 2005: *Language: A Biological Model*. Oxford: Oxford University Press.

- Millikan, R.G. 2009: Embedded Rationality. In M. Aydede and P. Robbins (eds.), *Cambridge Handbook of Situated Cognition*. Cambridge: University Press, 171-181
- Millikan, R.G. 2010 : On Knowing the Meaning; With a Coda on Swampman. *Mind* , 119 No. 473, 43-81.
- ONeill, D.K and A. Gopnik 1991: Young childrens ability to identify the sources of their beliefs. *Developmental Psychology*, 27, 390-397.
- Putnam, H. 1975a: *Mind, Language and Reality; Philosophical Papers I*. Cambridge UK: Cambridge University Press.
- Putnam H. 1975b: The meaning of meaning. In K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, vol. VII. Minneapolis: University of Minnesota Press. Also in Putnam 1975a.
- Whitcombe, E.L. and E.J. Robinson 2000: Childrens decisions about what to believe and their ability to report the source of their belief. *Cognitive Development* ,15.3, 329-346.
- Wimmer, H., G.J. Hogrefe and J. Perner 1988: Childrens understanding of informational access as a source of knowledge. *Child Development*, 59, 386-396.

Philosophy Department  
University of Connecticut

WORD COUNT 9741