chapter 9

Ruth Millikan

CHAPTER NINE
DIRECT PERCEPTION THROUGH LANGUAGE

I will argue that understanding language is simply another form of sensory perception of the world. I have already argued that perception is a way of understanding natural signs or, better, of translating natural signs into intentional signs. So this will help pave the way to the view that understanding language is very much like understanding natural signs.

A sign of a world affair that in turn signs a second world affair is itself a sign of that second affair (Chapter Four). Similarly, if the second sign is a sign of a third --and so forth. And there is always a direct semantic mapping function from the first sign to the last affair signified. A certain sound may signify that the dehumidifier has come on when heard from our bedroom at home, and this in turn may signify that the local power failure is over (a frequently recurrent sign in the rural area where we live). In our summer cabin an indistinguishable sound may signify that the refrigerator has come on, in turn signifying that we are not yet out of propane. To interpret these signs, you must be sensitive to the sign domains they inhabit. But in this case, the domain in which the sound signals the dehumidifier and the domain in which the dehumidifier signals the power are the same. Likewise, the domain in which the sound signifies the refrigerator and the refrigerator signals a non-empty propane tank are the same. A child then might simply hear the dehumidifier sound as the sound of electric power coming on, not being aware that it is the dehumidifier that produces what she reads as a sound of electric power, or the child might hear the sound of the refrigerator directly as a sign of propane hence that we won't have to go to town today for gas.

Similarly, suppose that certain patterns on the rabbit's retina are natural signs signifying presence of a fox and these patterns mediate between the fox and the rabbit's awareness of the fox. It will not be necessary that the rabbit possess an intentional representation of the retinal patterns in order to recognize the fox. Certain patterns traveling up its optic nerves will typically be natural signs of what is happening on its retina, but they will not be intentional signs of this. An intentional sign signifies only what it is used to represent when it operates normally. The optic nerve patterns are not used to guide either the rabbit's behavior toward its retina nor any inferences about its retina. It does not matter to the proper functioning of these patterns that they have been produced in a normal way from retinal images, rather than, say, by experimental electrodes, so long as they coincide with a fox. The images on its retina are of no more concern to the rabbit than the direction of magnetic north is to the anaerobic bacteria we encountered in Chapter Six, and for exactly the same reason. It is true, of course, that I can think about the images on my retina if I like, perhaps even know quite a lot about them. But when it comes to perceiving a fox, I have no more need to represent retinal images intentionally in the process than does the rabbit.

Exactly similarly, the words that the dog hears when it's master says "Go for a walk?" may move the dog directly to the expectation of a walk, that is, to an anticipatory mental representation of a walk, without the mediation of intentional representations of the master's intentions, or of the sounds it's master has made, or of the words it's master has spoken. You and I are capable of mentally representing those intentions, those sounds and those words, but the dog, probably, is not. Certainly he has no need to. The question naturally arises, then, whether it is necessary for you or me to harbor intentional representations of the sounds, the phonemes, and/or the words that mediate when someone calls out to us with the message, for example, "Dinner's ready!" Why should it be more complicated for us than for the dog?

Well, I think it is more complicated. First, let me explain why it has to be more complicated. Then let me explain why, despite these complications, there still is an important

sense in which, in routine cases, perception of the world though the medium of language can sensibly be called "direct perception."

Consider for a moment the connectionist network, VisNet, which was trained to recognize each of seven different faces, each presented through photographs at nine different angles in succession (using Hebbian induction and a trace rule), each succession of nine views of the same face being presented to the network 900 times. The result was that the network was able to recognize each of these faces with great reliability from each of these nine angles (McLeod et al 1998, p. 294 ff). This is quite an achievement for a contemporary connectionist net. But if such a net were now presented with an eighth face, it would take it just as long or longer to learn to recognize that new face as it took it to learn each of the seven previous ones, indeed, perhaps longer, because of interference from traces already laid down. A face recognizer built like VisNet would be like a brain that relied on semantic mapping functions that went directly from retinal stimulations to inner intentional signs of, say, Johnny, without first passing through a stage that represents the objective shape of Johnny's face and then interpreting this shape as a sign of Johnny (chapter Four). Or it is like the child who relies on a semantic mapping function that goes directly from hearing the words "president of the paleontological society" to an inner sign for Mommy, without first going through a stage at which the words "president," "paleontological" and "society" are understood (Chapter Four). Or, of course, it is like the dog who goes directly from hearing the sounds "Go for a walk?" to expecting a walk without going through the process of first recognizing the phonemes, then the words these represent, then the syntactic structures involved, moving finally to an interpretation of what they represent.

If you have need to learn many new faces quickly, simulating VisNet is not a good way to go about it. It is more efficient if you first put in place the ability to recognize <u>any</u> shape, or at least any face-like shape, as such, from <u>any</u> arbitrary angle and at <u>any</u> arbitrary distance. It is better if you first put in place the capacity called "perception of shape constancy." Once you have the general capacity to recognize same-shape-again, for pretty much any shape, from pretty much any perspective, learning to recognize Johnny involves only learning which face-shape is a recurrent sign, in your locale, of Johnny. Learning to recognize Sally uses the same general capacities over again, with the minor variation that you must learn which face-shape is a local sign of Sally. Similarly, if only the dog could learn first to recognize each of the various phonemes in his master's language through the variety of their possible acoustical manifestations, then learn to recognize the words that the presence of various strings of these phonemes may express, then learn to recognize the syntactic forms that can be expressed by various strings of these words, then add a grasp of the semantic mapping functions from words-plus-syntax in his master's language onto what these sentences conventionally represent, it would not take him dozens and dozens of exposures to learn the significance of each new one of the trivial number of expressions of which he may (roughly) grasp the meaning by the end of his life.

(I have spoken here of the phonemes as <u>expressing</u> words, not as composing words, and of strings of words as <u>expressing</u> syntactic forms because words considered <u>as</u> signs do not equal strings of phonemes. Exactly the same phoneme strings can compose different words. Words and syntactic forms are (aspects of complete) signs and, as we know, exactly the same physical type, depending on the domain it is in, can compose quite different signs, signs of quite different things.)

That human speech perception is routed through the recognition of phonemes and only later of words is evidenced, for example, by the fact that all same-sounding words are primed when a word is presented, even if they have quite different meanings (Swinney 1979). Words must be

recognized by context after recognizing the phonemes expressing them. And the same is often true of syntactic forms. Just the parts and ordering of parts of which they are composed, their surface forms, do not identify them. Context, often including inner context (the particular words within the syntactic arrangement), is required to keep track of the meme families that are syntactic forms. This is evident in examples of syntactic ambiguity and mood ambiguity, for example, in the sign "Recycle cans and waste paper" and in the sentence, "You will attend university," which might be descriptive (a fortune teller's prediction) or might be directive (Father's orders).[1]

There seems to be a good reason, then, why humans, in understanding human language forms, should go through a process in which intentional representations of things that are signs of signs and so forth are formed, passing through a number of layered stages of intentional representation in the process of translating public language signs into inner representations of world affairs. Similarly, neurological evidence suggests that ordinary visual perception involves the translation of gradients of luminance across the retina as signs of various rudiments of visual form, such as lines or edges with a particular orientation, right angles, ocular disparity, directional movement, color edges, and so forth, which are then interpreted as signs of such constancies as shape, mentioned above, and of a variety of other property constancies, such as constancies for color, size, texture, quality and direction of movement, and so forth.[2] Taking another example from neurology, Zipser and Anderson (1988) have developed an extremely plausible connectionist model of the way representations of the direction of light points relative to the eye are combined with representations of the orientation of the eye within the head to yield representations of the orientation of the light points relative to the head, and they have demonstrated a very close match to the activation profiles actually found among single neurons in the posterior parietal cortex apparently responding to light-head angles. That is, the evidence is that the brain reads signs of light-eye relations coupled with signs of eye-head relations as signs of light-head relations in a systematic way. It does not learn to recognize each light-head relation separately. It uses representations at various levels, and translates from one level to the next systematically.

Why is this kind of arrangement, then, not an example, exactly, of "indirect perception"? Why isn't the obvious conclusion to be drawn that even the most basic representational functions of the brain involve it in making inferences --inferences prior to the perception of objects in space and prior to the perception of objects represented through language? It is easy to imagine falling into a verbal dispute over this issue. What I will do is to argue that there are at least three important differences between the way the brains reads natural signs and signs of signs, and so forth, of the world, and the way indirect perception was traditionally understood to go. The result, I hope, will be to convince the reader that the terms "direct" and "indirect" as they were traditionally understood, fail to have any useful application in the realm of perception. Then I will introduce my own suggestion about a sensible use of the terms "direct" and "indirect."

Traditional theories of "indirect perception" contrasted the perception of objects in the outer world with a different kind of perception which they took to be direct perception, namely, perception of sense impressions, or sense data or sensations. The idea was that these sense impressions were the first objects intentionally represented by the mind, and that the mind had to perform inferences in order to move to representations of anything outside. Perception of impressions was typically assumed to be not only direct but also infallible. One couldn't represent one's own sense impressions to oneself wrongly. Perception of outside objects, on the other hand, was fallible, subject to illusion. But on the description of perception of the outer world we have

been considering, no such contrast can be drawn. The mind/brain does not begin by representing either happenings on the sensory surfaces, like light gradients across the retina, or inner things, like visual impressions. The first steps in perception involve reacting to natural signs of features of the outer objective world by translating them into inner intentional representations of those outer features, for example of edges, lines, angles of light sources in relation to the eye, and so forth. These are outer things, or relations to outer things, not inner objects. Nor, of course, does the brain represent nor the mind become aware of the vehicles in the brain that do the representing. It does not represent its own representations during perception. If direct perception has to be of inner things, there simply is no direct perception, at least none that is involved during the process of perceiving he outer world . Moreover, on every level, inner intentional representations are fallible. Nothing is direct in the sense of being epistemically "given" to mind. Interpretation of signs is always fallible, chancy, in very principle.

The perception of outer-world objects is not "indirect" either, not in the classical sense. For the traditional view of "indirect perception" was that representations of the outer world were derived by the use of inference. I have described the derivation via perception of a representation of, say, Johnny, or of the affair represented by a sentence one hears, as involving a series of fallible translations from one sign or signs into others. Taking an affair as a sign of another affair is reacting to the first affair by translating it into an intentional representation of the other. It doesn't matter whether the first affair, the one taken to be a sign, is a natural sign or an intentional sign. In either case, sign interpretation is just translation. Where the first sign is an intentional sign, this can be put by saying that interpreting it is reacting to the sign vehicle by translating it into or deriving from it another sign vehicle. The question that arises, then, is whether there is a significant difference between translation and inference. For inference also seems to be reacting to one sign vehicle (or perhaps a pair) by translating what it (or they) signify, or at least a portion of this signification, into another sign vehicle.

Here are two differences between the interpretation of signs, and of signs of signs, postulated to occur within perceptual processes, and the traditional paradigm of inference, a movement from beliefs to further beliefs. First, translations cannot easily be modeled after any of the recognized forms of inference, deductive, inductive or abductive. This is because they don't seem to have major premises, either explicit or suppressed. Consider the move from seeing the shape that is like that of Johnny's face to representing the presence of Johnny. The premise "all shapes like that coincide with presences of Johnny" is not presupposed. It may very well not be true, in fact, and the perceiver may be quite prepared to discover that it is not. Somewhere in the world is another boy whom one could not tell from Johnny by a glance at his face. Nor is the premise "most face shapes like that coincide with presences of Johnny" presupposed. What is relied on here is not an implicit major premise, but a fallible ability you have to track a kind of sign domain. The capacity to track, or the luck to happen to remain within the domain of a certain local sign, is not helpfully modeled as a major premiss, any more than your luck in not unexpectedly hitting the wrong kind of patch of ice or sand when trying to keep upright on a bicycle would be helpfully modeled as a major premise.[3]

And there is another difference between paradigm inference and the transitions from sign to sign that take place during perceptual processing. Primitive inner signs, such as the sign indicating the angle of the light point relative to the eye, are dedicated intentional signs. They are likely to have just one or two jobs to do, for example, interacting with signs of the position of the eye in the

head to produce signs of of the relation of the light to the head. That is all. The classical model for inferences, on the other hand, comes from the realm of belief. It is assumed that beliefs form a system of representations such that any representation in that system could, in principle, interact with any other, either directly or through intermediate beliefs. None are isolated, nor are any groups of them isolated from the main body of beliefs. Further, their interactions do not take the form of set algorithms, but may branch in any of numerous directions from a given starting point or points. Beliefs are not, as such, dedicated in advance to any particular purposes. They may help to serve any of a wide variety of special purposes, not determined for them in advance by the systems that have fashioned them. They are designed to interact with other beliefs and desires in a flexible manner, acquiring more specialized purposes depending on their inner intentional environments.

For these reasons I think that assimilating the translations that take place during ordinary perception of objects to inferences will mislead rather than enlighten. There is no such thing as "direct perception," in the sense originally intended by those claiming that ordinary perception was "indirect," and there is no such thing as "indirect perception," in that sense, either. If a distinction is to be drawn then between arriving at a representation of the outer world "directly" versus "indirectly," it will have to be drawn in another way. Perhaps it would be best to drop the terms "direct" and "indirect" altogether in this context and stick to describing "just the facts, Mam, just the facts." On the other hand, it is useful to have a term to describe moves from sensory inputs to representations that do not pass through stages sensibly called "inference" but involve at most only translations. I propose to save the term "direct perception" for this purpose.

My job now is to argue that coming to believe something by being told it is so, in the typical case, is the formation of a direct perceptual belief. Forming a belief about where Johnny is on the basis of being told where he is, is just as direct a process (and just as indirect) as forming a belief about where Johnny is on the basis of seeing him there.

You and I have concepts of edges, of various shapes, of phonemes, of words and of sentences. We can use these concepts in developing theories, for example, of how objects and properties are perceived and of language perception and production. Very small children do not have concepts of such things as edges, shapes, phonemes, words and sentences. For example, as mentioned in Chapter Seven, a small child will tell you that "ghost" is not a word because there aren't any ghosts. Most children are unable to segment words into their component phonemes or to recognize phonemes as recurring entities until about five or six years (Liberman et al 1974). Certainly they have no concepts of phonemes as such. But that their brains employ representations of edges, shapes, phonemes and words by the time they are beginning to talk seems unquestionable. Their ability to recognize new faces and to understand what is said to them clearly depends on these capacities. This is also true, it appears, for morphemes. Children will generalize from reading, say, the "er" in "higher" and "bigger" to reading "smaller" and "smarter" but not to "weather" and "bicker" (B. Byrne 1996) Yet probably even most adults do not have concepts of morphemes. Not every capacity to produce intentional signs to be used on the way to producing further intentional signs has the right use or a general enough use plausibly to be considered a "concept."

There is evidence that when we hear someone speak, normally what is said goes directly into belief, exactly as when we observe some event happening directly (Gilbert 1993). We do not first understand what is said and then evaluate whether to believe it. Rather, we first believe what is said and then, if we are not under too much cognitive stress, we may think it over critically and

reject it. Subjects who are under too much cognitive load, say, they are trying at the same time to count backwards from 1000 by threes, strongly tend simply to believe whatever they hear. In general, there seems to be no reason to suppose that there is only one particular level of distality at which each sensory modality perceives "directly," in the sense I have defined.[4] When you watch television, you usually directly see what is depicted. You see the newscaster's face and what he is wearing and you hear what he is saying. But if you change your frame of mind slightly, say, you are wondering whether to purchase this TV set or not, you may stop seeing and hearing at this level and instead concentrate only on the quality of the reception. Or you may see the pixels on the screen flashing in patterns, especially if the reception is not very good, or if you are a repair man trying to diagnose the set's maladies. Similarly, as an adult, you can directly perceive the phonemes or directly perceive the words being uttered by a speaker if you want to. But usually you perceive only the world affairs spoken about. Just as what we see is dependent on the depth at which we focus our eyes, the distality of what we directly see or hear depends on where we focus our minds.

Depending on the external media through which information is transmitted for perception, the very same world affairs may appear to the same sensory organs in different guises. Although we have surprisingly good color constancy perception under a variety of different lighting conditions, colors that are perceived as the same objective color do not have the same appearance under all conditions. Nor do shapes that are perceived to be the same objective shape have the same appearance from all angles. Ringing bells and clacking sticks are easily recognized for what they are whether heard through air, under water, in a sound absorbing chamber, or in an echo chamber, but they do not sound the same under these different conditions. Rain does not sound the same when heard falling on the roof, on earth, on snow, and on the water, even though it may be directly perceived as rain through any of these media. Exactly similarly, rain has a different sound when the medium of transmission is the English language ("It's raining!"). And it sounds different again when the medium of transmission is French or German. What world affairs sound like when transmitted through language depends on the language community you are in.

Why is the notion so unintuitive that understanding and believing what is said to you is just another level of natural sign reading and, often, just another form of direct perception? One reason is that what is given to us in ordinary perception is always given as in some rather definite current relation to us. It is given as happening at the time we perceive it, as happening relatively nearby, and often as bearing quite an exact spatial relation to us. This kind of information is needed to guide action, for how one can presently act on a thing always depends on its present relation to one. Ordinary perception is for immediate action, whereas what one learns through language is not typically used that way. Usually I am not told what exact spatial and temporal relations the world affairs being presented through language have to me here and now. Let us take a careful look at this difference between ordinary perception and perception through language.

One of the many traits that seem to distinguish us rather sharply from other species is an enormous flexibility in learning to read new recurrent signs of affairs at different levels of distality and defined by different kinds of mapping functions. For example, learning to comb hair in a mirror is very easy for us, but interpreting what is in mirrors is not possible at all for most other animals. Some have thought that this has something to do with the development of a "self concept," but there is no coherent argument for this. Why should seeing a part of one's body through a mirror and correctly interpreting it in the way necessary to guide one's reaching and touching behavior

require a self concept, if seeing a part of one's body in the normal way for the same purpose does not? Nor should it be thought that mentally representing the relation of something to oneself, perceiving the relation of something to oneself, requires that one represent oneself explicitly. Recall, for example, that bees represent the relation of nectar to hive and sun without explicitly representing either the hive or the sun (Chapter Seven). What is required in using a mirror is only that one accommodate governance of one's perceptions and guided motions to a new semantic mapping function in taking account of to the relation of seen objects to oneself. In the rearview mirror, I directly see that there is a car behind me. The car behind guides my motion in relation to it appropriately and directly. Only a very few of the higher primates are capable of making this shift to seeing things in mirrors.[5] For example, as mentioned in Chapter Four, when a kitten first sees itself in the mirror, it tries to approach the other kitten, to smell it and touch it. Failing in this, it then tries to look behind the mirror. Finding nothing there, it immediately adopts the attitude that mirrors produce merely holes in information space. You cannot see anything through a mirror any more than though murky water. It is quite impossible to interest a kitten in its reflection after this first disappointing encounter.

Consider what is required to understand what a photograph represents. There is some question to what degree animals other than humans can learn to "see" anything at all in photographs. Pigeons can be taught to sort photographs into those that picture trees versus people versus water and so forth (Herrnstein 1976). They are rewarded for doing this, of course, so the mental representations they derive from the pictures have a function. And it is very likely that their capacity to do this kind of sorting rides piggyback on their prior capacities to recognize actual trees, people and water. But it seems out of the question that they actually acquire any information from these pictures about what is pictured. This is because a photograph contains no information about the relation to the current observer of what it depicts. And what use would a pigeon have for a mental representation, derived from a photograph, that there once existed, sometime and somewhere or other, a tree looking just so in front of a house looking just so? Similarly, just as a kitten soon stops bothering with mirrors, most animals soon stop looking at television. If they don't stop looking, it is likely that they misinterpret what they see, as did the small daughter of a colleague of mine who asked "Daddy, how did you get in there?" after watching her father on local TV.

That human perceivers can retrieve information from photographs and television depends on their capacity to use information about distal affairs that are not represented or yet understood as having definite and useful relations to themselves. Exactly similarly, information presented through human language forms does not typically include information about the relations to the hearer of the affairs presented. The capacities required to understand human language include, then, not only a marvelous flexibility in accommodating to new semantic mapping functions, but also the capacity mentally to represent, which requires having some use for (Chapter 6), information that does not include the relations to you of the things the information is about. But if you are willing to extend notions of perception far enough to cover "seeing" the newscaster on TV, the extension to "hearing" the news events through his recounting is the same kind of extension. There is no shift in directness of perception, but only a lessening of content in what is perceived. Information about relations to self have dropped out.

But there is another important reason too why believing what is said to you may not seem to resemble direct perception. The reliability seems to be quite different. Recall the discussion of verbs of perception in Chapter Five. These verbs equivocate, posing sometimes as achievement

verbs and other times as verbs merely of aiming or trying. This is one of the factors that may make it seem that ordinary perception of objects and affairs in the world is far more reliable than acquiring information through language. You cannot see what isn't there, but you certainly can hear what isn't true. But, of course, you also can see or hear what isn't there --"hearing voices," for example-- just as you can hear what isn't true. When we speak of "hearing" that such and such is the case through the medium of language, however, the aiming sense of "hearing" predominates whereas ordinary "hearing," like "seeing," is more likely to be meant in an achievement way. "Yesterday upon the stair I saw a man who wasn't there"[6] has a very peculiar ring. So does "Yesterday upon the stair I heard a man who wasn't there." But "Yesterday I heard there was a man on the stair but there was no man there" is quite straight forward.

There is a reason for this. Ordinary perception is indeed considerably more reliable than what one hears said, at least under common circumstances. It is not easy to fool ordinary perception. To create strong perceptual illusions generally requires a good deal of knowledge about the perceptual mechanisms and often quite special equipment, for example, of the kind optometrists have in their examination rooms. But surely these illusions should not be classed as indirect perceptions just because they are deceiving. There is no such thing as infallible perception of anything. Suppose you have new lenses with a new strong correction for astigmatism so that the sidewalk in front of you looks curved or wavy? Do you start perceiving or "observing" the world "directly" again only after adjusting to the glasses? It is reasonable to say that you see the news commentator on television, but what if the film strip he shows you is dubbed or outright faked? Dubbing of films is currently the rule rather than the exception. Is there a difference of kind between believing what you apparently see when a film as been dubbed and believing what you hear someone say when it's false? In the modern world, if you want to believe only what's true, you often have to apply heavy filters to other methods of perception as well as to perception through language.

The picture I want to leave you with, then, is that coming to believe, say, that Johnny has come in by seeing that he has come in, by hearing by his voice that he has come in, and by hearing someone say "Johnny has come in," are normally equivalent in directness of psychological processing. There is no reason to suppose that any of these ways of gaining the information that Johnny has come in requires that one perform inferences. On the other hand, in all these cases it is likely that at least some prior dedicated representations must be formed. Translations from more primitive representations and combinations of these will be involved. If one insists on treating all translation as a form of inference, then all these require inference equally. In either event, there is no significant difference in directness among them.

FOOTNOTES

1. For more detail on this point, see Millikan 1984, Chapter 3. What I call "memes" in this essay are there called "members of first order reproductively established families." The relevant sense of "word," in the present context, is what was there called a "least type" of word.

2. See, for example, Norman 2000, but the general idea here is ubiquitous in the neurological literature on perception.

3. Another possibility might be to try to model the major premiss behind the translation of one sign into another as an implicit identity judgment. "That shape there IS Johnny's face." That there actually are no such things as identity judgments, that is, that identity sentences don't correspond to any mental representations, is argued in (Millikan 2000) Chapter 12. Identity sentences are an example of language forms that have truth conditions but whose functions are not to cause mental representations that have truth conditions.

4. I am using "distal" here as it was used in Chapter Four to refer to the length of a chain of signs of affairs that are themselves signs.

5. Interestingly, according to Epstein, Lanza and Skinner (1981), pigeons are able to locate spots on their bodies using a mirror. The authors take this to be proof that recognizing parts of one's body in a mirror does not require a self concept.

6. The verse continues, "He wasn't there again today; My God I wish he'd go away!"