Chapter 3
Representational Systems

Some behavior is the expression of intelligent thought and purpose. Clyde goes to the kitchen because he wants another beer and thinks there is one left in the refrigerator. Whether or not they are *causes* of behavior, Clyde's reasons—his desire for a beer and his belief that there is one in the fridge—are certainly thought to *explain* his behavior. They tell us *why* he made the trip to the kitchen.

This is our ordinary way of explaining behavior (at least, those behaviors we think of as purposeful). It is so familiar, so utterly commonplace to all of us, that it is hard to see how there can be a problem with this type of explanation.

There is, nonetheless, a problem in understanding how this familiar pattern of explanation can take—or hold—its place alongside the emerging neuroscientific picture of living organisms. How do, how *can*, thoughts and purposes determine what we do when what we do, at least what our bodies do, seems so completely dependent on, and therefore determined by, those neuronal processes and mechanisms described, in increasingly rich detail, by neurophysiologists? If the neurophysiologists don't invoke thoughts, purposes, intentions, desires, hopes, and fears to explain the behavior of a person's body, what excuse (besides ignorance) do *we* have for appealing to such notions to explain the behavior of the person?

We have already taken the first step toward a better understanding of this apparent conflict. The first step is to understand the difference between a person's behavior and whatever bodily movements and changes constitute this behavior. An understanding of the difference between Clyde's going to the kitchen and the movements that get him to the kitchen is essential to an understanding of why an explanation of the one is not an explanation of the other. Knowing why Clyde went to the kitchen isn't the same as knowing why his legs moved so as to bring him into the kitchen; and knowing the causes of limb movement, at whatever level of biological detail, is not the same as knowing why he went to the kitchen. These are different explanatory games. Our familiar way of explaining purposive behavior in terms of an agent's intentions and beliefs does not *compete*

with a neurobiological account of muscular activity and, hence, with a mechanistic account of motor output. It is, rather, an attempt to explain something altogether different: *behavior*, not output.

There is, however, a second step that must be taken. As yet we have no idea of how ordinary explanations, explanations couched in terms of an agent's *reasons*, explain. Since behavior has been identified with a process, with one thing's *causing* another, are reasons supposed to be the cause of one thing's causing another? If so, how is this supposed to work, and what is it about reasons that gives them this peculiar efficacy?

In order to answer these questions, in order to take this second step, it will be necessary to spend some time examining the idea of a representation. For beliefs, normally a prominent part of one's reasons for acting (desire being another prominent part), are special kinds of representations. Beliefs are those representations whose causal role in the production of output is determined by their meaning or content—by *the way* they represent what they represent. The general idea of a representational system is examined in this chapter. The special topic of belief is reserved for chapter 4.

3.1 Conventional Systems of Representation: Type I

By a representational system (RS) I shall mean any system whose function it is to indicate how things stand with respect to some other object, condition, or magnitude. If RS's function is to indicate whether *O* is in condition *A* or *B*, for instance, and the way RS performs this function (*when* it performs it) is by occupying one of two possible states, *a* (indicating that *O* is *A*) and *b* (indicating that *O* is *B*), then *a* and *b* are the expressive elements of RS and *what they represent* (about *O*) is *that* it is *A* (in the case of *a*) and *that* it is *B* (in the case of *b*).

Depending on the kind of function involved, and on the way a system manages to carry out this function (the way it manages to *indicate*), representational systems can be variously classified. What follows is one possible classification. My chief interest is in *natural* representations (systems of Type III), but the special properties of such systems are best understood by comparing and contrasting them with their conventional (to varying degrees) cousins. So I begin with conventional systems of representation.

Let this dime on the table be Oscar Robertson, let this nickle (heads uppermost) be Kareem Abdul-Jabbar, and let this nickle (tails uppermost) be the opposing center. These pieces of popcorn are the other players, and this glass is the basket. With this bit of stage setting I can now, by moving coins and popcorn around on the table, represent the positions and move-

ments of these players. I can use these objects to describe a basketball play I once witnessed.

If memory fails me, I may end up misrepresenting things. I may move pieces of popcorn here when the players went there. The coins and the popcorn have been assigned a temporary function, the function of *indicating* (by *their* positions and movement) the relative positions and movements of certain players during a particular game. But these elements, the coins and the popcorn, obviously enjoy no intrinsic power to do what they have been assigned the function of doing—*indicating* the positions and the movements of various players in a game long since over. Whatever success they enjoy in the performance of their job obviously derives *from me*, from my memory of the game being represented and my skill in translating that knowledge into the chosen idiom. The popcorn and the coins indicate, and in this sense perform their assigned function, only insofar as *I* am a reliable conduit for information about the situation being represented and a reliable and well-intentioned manipulator of the expressive medium.

The coins and the popcorn do their job, then, only insofar as some *other* indicator system is functioning satisfactorily, only insofar as there is something in the manipulator of these symbols (in this case, something *in me*) that indicates how things stood on the basketball court at the time in question. If I am ignorant of what Oscar and Kareem did with the ball, the coins and the popcorn are unable to perform the function they have been assigned—unable to indicate, by their various positions and movements, what took place on the court that day. This is merely to acknowledge that these objects are, considered by themselves, representationally lifeless. They are merely my representational instruments.

The elements of Type I systems have no *intrinsic* powers of representation—no power that is not derived from us, their creators and users.[1] Both their function (what they, when suitably deployed, are *supposed* to indicate) and their power to perform that function (their success in indicating what it is their function to indicate) are derived from another source: human agents with communicative purposes. Many familiar RSs are like this: maps, diagrams, certain road signs (of the informational variety), prearranged signals, musical notation, gestures, codes, and (to some degree, at least) natural language. I call the representational elements of such systems *symbols*. Symbols are, either explicitly or implicitly, *assigned* indicator functions, functions that they have no intrinsic power to perform. *We*

1. That is, no intrinsic power to indicate *what it is their (assigned) function to indicate*. They may, of course, indicate something *else* in a way that is not dependent on us. For instance, the coins, being metal, indicate (by their volume) the temperature. They *could*, therefore, be used as crude thermometers. But, according to the story I am telling, this isn't their (assigned) function. If it was, then we would be talking about an RS of Type II.

give them their functions, and *we* (when it suits our purposes) see to it that they are *used* in accordance with this function. Such representational systems are, in this sense, *doubly* conventional: *we* give them a job to do, and then *we* do it for them.

3.2 Natural Signs and Information

In contrast with the relationship between popcorn and professional basketball players, we don't have to *let* tracks in the snow, bird songs, fingerprints, and cloud formations stand for the things we take them to indicate. There is a sense in which, whether we like it or not, these tracks, prints, songs, and formations indicate what they do quite independent of us, of how we exploit them for investigative purposes, and of whether we even recognize their significance at all. These are what are sometimes called *natural signs*: events and conditions that derive their indicative powers, not (as in the case of symbols) from us, from our *use* of them to indicate, but from the way they are objectively related to the conditions they signify.

To understand conventional systems of representation of Type II and the way they differ from RSs of Type I, it is important to understand the difference between symbols and signs. In systems of Type II, natural signs are *used* in a way that exploits their *natural* meaning, their *unconventional* powers of indication, for representational, and partly conventional, purposes. This makes systems of Type II a curious blend of the conventional and the natural. It is the purpose of this section to say something useful about signs and their meaning in preparation for the description of representational systems of Type II. This, in turn, will prepare the way for our discussion of the representational systems that are of real interest to this project: natural systems of representation.

Although a great deal of intelligent thought and purpose went into the design and manufacture of an ordinary bathroom scale, once the scale has been finished and placed into use there is nothing conventional, purposeful, or intelligent about its operation. This device indicates what it does without any cooperation or help from either its maker or its user. All you do is get *on* it. It then gives you the bad news. Somebody put the numbers on the dial, of course, and did so with certain intentions and purposes; but this is merely a convenience, something that (to use fashionable jargon) makes it user-friendly. It has nothing to do with what the instrument indicates. A clock doesn't stop keeping time if the numerals are removed from its face. The symbols on a clock or on a bathroom scale merely make it easier for us to *tell* what the pointer positions *mean*. They do not change what these pointer positions indicate.

The same is true of any measuring instrument. As long as an instrument is connected properly and functioning normally, it behaves in accordance

with electrical and mechanical laws whose validity is quite independent of its creator's or its user's purposes or knowledge. Furthermore, these laws, by determining whether and (if so) how the pointer positions are correlated with weights, times, pressures, and speeds, determine what these pointer positions indicate about weights, times, pressures, and speeds.

Some people think that all indication is indication *for* or *to* someone. Gauge readings and naturally occurring signs (e.g., tracks in the snow) do not indicate anything if there is no one *to whom* or *for whom* they do this. Gauge readings are like trees falling in the forest: if no one is around to hear, there is no sound; if no one peeks at the scale, it doesn't indicate anything about anyone's weight. Tracks in the snow, fingerprints on a gun, and melting ice do not indicate anything about the animals in the woods, the person who touched the gun, or the temperature *unless* someone observes the tracks, the prints, or the melting ice and makes an appropriate inference. If no one knows that quail, and *only* quail, make tracks of *that* kind, then, despite this regularity, the tracks do not indicate that there are (or were) quail in the woods.

This view, I submit, is merely a special version of the more general and even more implausible idea that nothing is true unless it is true for someone, unless someone knows (or at least believes) it. I do not intend to quarrel about this matter. I shall simply assume that if one mistakes a perfectly reliable and properly functioning boiler-pressure gauge for something else, thinks it is broken, completely ignores it, or never even sees it—if, in other words, the registration of this gauge does not indicate what the boiler pressure is *to anyone*—it nonetheless still indicates what the boiler pressure is. It just doesn't indicate it *to* anyone. And, for the same reason, if, for superstitious reasons, everyone takes the color of the wooly caterpillar's fur as a indication or sign of a cold winter, everyone is simply wrong. That isn't what it means. Taking something to be so, taking it to be not so, or not taking it to be either does not make it so, does not make it not so, and does not make it neither. And this holds for what things indicate as well as for where things are and what they are doing.

I have occasionally used the verb "mean" as a synonym for "indicate." Let me explain. Paul Grice (1957) distinguished what he called a natural sense from a non-natural sense of the word "meaning." The natural sense of "meaning" is virtually identical to that of "indicate," and that is how I shall normally use the word. The 24 rings in a tree stump, the so-called growth rings, mean (indicate) that the tree is 24 years old. A ringing bell—a ringing *door*bell—means (indicates) that someone is at the door. A scar on a twig, easily identified as a leaf scar, means, in this natural sense, that a leaf grew there. As Grice observes, nothing can mean that *P* in the *natural* sense of meaning if *P* is not the case. This distinguishes it from non-natural meaning, where something (e.g., a statement) can mean that *P* without *P*'s

being the case. A person can *say*, and *mean*, that a quail was here without a quail's having been here. But the tracks in the snow cannot mean (in this natural sense of "meaning") that a quail was here unless, in fact, a quail *was* here. If the tracks were left by a pheasant, then the tracks might, depending on how distinctive they are, mean that a pheasant was here. But they certainly do not mean that a quail was here, and the fact that a Boy Scout *takes* them to mean that cannot *make* them mean that.

Furthermore, even if *P* does obtain, the indicator or sign does not mean (indicate) that *P* is the case unless the requisite *dependency* exists between the sign and *P*. Even if the tracks in the snow *were* left by a quail, the tracks may not mean or indicate that this is so. If pheasants, also in the woods, leave the very same kind of tracks, then the tracks, though made by a quail, do not indicate that it was a quail that made them. A picture of a person, taken from the back at a great distance, does not indicate *who* the picture is a picture of if other people look the same from that angle and distance.

If a fuel gauge is broken (stuck, say, at "half full"), it *never* indicates anything about the gasoline in the tank. Even if the tank *is* half full, and even if the driver, unaware of the broken gauge, comes to believe (correctly, as it turns out) that the tank is half full, the reading is not a sign—does not mean or indicate—that the tank is half full. Broken clocks are *never* right, not even twice a day, if being right requires them to *indicate* the correct time of day.

When there is any chance of confusing this use of the word "meaning" with what Grice calls non-natural meaning—the kind of meaning associated with language, the kind of meaning that is (I shall later argue) closer to what it is the *function* of something to mean (naturally) or indicate—I shall either combine the word "meaning" with the word "natural" or use it together with its synonym "indicate." The word "represent" is sometimes used in a way that I am using "indicate" and "mean" (naturally). Since I wish to reserve the idea of representation for something that is closer to genuine meaning, the kind of meaning (Grice's non-natural meaning) in which something can mean that *P without P*'s being the case, I will *never* use the words "represent" and "indicate" interchangeably. As I am using these words, there can be no *mis*indication, only misrepresentation.

The power of signs to mean or indicate something derives from the way they are related to what they indicate or mean. The red spots all over Tommy's face mean that he has the measles, not simply because he *has* the measles, but because people without the measles don't have spots of that kind. In most cases the underlying relations are causal or lawful in character. There is, then, a lawful dependency between the indicat*or* and the indicate*d*, a dependency that we normally express by conditionals in the subjunctive mood: if Tommy didn't have the measles, he wouldn't have those red spots all over his face. Sometimes, however, the dependency

between a natural sign and its meaning derives, at least in part, from other sources. It is partly the fact, presumably not itself a physical law, that animals do not regularly depress doorbuttons while foraging for food that makes a ringing doorbell *mean* that some *person* is at the door. If squirrels changed their habits (because, say, doorbuttons were made out of nuts), then a ringing doorbell would no longer mean what it now does. But as things *now* stand, we can say that the bell would not be ringing unless someone was at the door. It therefore indicates or means that someone is at the door. But this subjunctively expressed dependency between the ringing bell and someone's presence at the door, though not a coincidence, is not grounded in natural law either. There are surely no laws of nature that prevent small animals from pressing, or randomly falling meteorites from hitting, doorbuttons. There certainly is nothing in the laws of physics that prevents an occasional short circuit in the electrical wiring, something that might cause the bell to ring when no one was at the door. Normally, though, these things don't happen. At least they have never happened to *me*. And this is no lucky coincidence, no freaky piece of good fortune. It isn't like getting a long run of heads while flipping a (fair) coin. Chance correlations between two variables, no matter how prolonged, are not enough. In order for one thing to indicate something about another, the dependencies must be genuine. There must actually be some condition, lawful or otherwise, that *explains* the persistence of the correlation. This is the difference between a lucky run of heads obtained with a fair coin and the not-at-all-lucky run of rings when someone has been at my door, a difference that enables my bell (but not coin flips) to indicate something about the correlated condition. This, of course, is a fact about *my* house, *my* neighborhood, and *my* doorbell wiring. If your house or neighborhood is different, maybe the ringing of *your* doorbell means something different.[2]

In many cases of biological interest, a sign—some internal indicator on which an animal relies to locate and identify, say, food—will only have this kind of local validity. It will, that is, be a reliable indicator only *in* the animal's natural habitat or in conditions that approximate that habitat. Flies,

2. Fodor (1987b) mentions an interesting illustration of this phenomenon discussed by David Marr and his associates: an algorithm (in the perceptual system) for computing three-dimensional form from two-dimensional rotation. The algorithm is not strictly valid, since there are worlds in which it reaches *false* three-dimensional conclusions from *true* two-dimensional premises—worlds in which spatial rotations are not rigid. Nevertheless, the algorithm is truth-preserving in the circumstances in which it is in fact employed—viz., *here*, in our world. Add to this the fact that the perceptual mechanisms that exploit this algorithm were evolved *here*, in *this* world, and we have a biological example of a uniformity—not lawful, but not fortuitous either—that enables sensory "premises" about two-dimensional rotations (that is, premises describing the two-dimensional transformations of the retinal image) to indicate something about the three-dimensional world we live in.

for instance, when given a choice between nutritionally worthless sugar fructose and some nutritive substance like sorbitrol, will invariably choose the nutritionally worthless substance and starve to death. Surprising? Not really. Under *natural* conditions (Grier 1984, p. 536) the substances that stimulate the receptors *are* nutritional. Under natural conditions, in a fly's normal habitat, then, receptor activity indicates a nutritional substance. Furthermore, the correlation between receptor activity and nutritional value of its activator is no accident. There is something that explains it. Flies would not have developed (or maintained without modification) such a receptor system in environments where such a correlation did not exist. The same is true of me and my doorbell. I would not keep a doorbell system that did not convey the desired information, that did not (because of pesky squirrels, say) indicate what it was installed to indicate. I would, as I assume the flies (over many generations) would, get a more discriminating detector.

I have elsewhere (1981, 1983), under the rubric *information*, tried to say something more systematic about the idea of an objective, mind-independent, indicator relation. Aside from the above brief remarks tracing the idea of natural meaning to the objective relations of dependency between a natural sign and its meaning, between the indicator and what it indicates, I will not here attempt to recapitulate that earlier analysis. Nor will I presuppose the details. Sufficient unto present purposes is the assumption—an altogether plausible assumption, I hope—that there is something *in* nature (not merely in the minds that struggle to comprehend nature), some objective, observer-independent fact or set of facts, that forms the basis of one thing's meaning or indicating something about another.[3] In what follows I shall occasionally, partly as a terminological convenience but also partly to exhibit the deep connections between representational systems and information-processing models of human cognition, advert to the idea of information. Talking about information is yet a third way of

3. This is not to say that descriptions of what something means or indicates are always free of subjective factors. We often describe what something means or indicates in a way that reflects what we already *know* about the possibilities. If there are only two switches controlling a light, the light indicates that one of the two switches is closed. Knowing, however, that *this switch* (one of the two) *isn't* closed, I take the light's being on as an indication that *the other switch* is closed. In this case, the light (is said) to indicate something that it would not indicate unless I, the speaker, *knew* something about other possibilities.

In this sense the meanings we ascribe to signs is relative. It is relative to what the speaker already knows about possible alternatives. This, however, doesn't mean that natural meaning is *subjective*. A person's weight isn't subjective just because it is relative, just because people weigh less on the moon than they do on earth. If nobody knew anything, things would still indicate other things. They just wouldn't indicate the specific sort of thing (e.g., the other switch is closed) we now describe them as indicating.

talking about the fundamentally important relation of indication or natural meaning. So, for example, if *S* (sign, signal), by being *a*, indicates or means that *O* is *A*, then *S* (or, more precisely, *S*'s being *a*) carries the information that *O* is *A*. What an event or condition (whether we think of it as a signal or not is irrelevant) indicates or means about another situation is the information it carries about that other situation.

3.3 Conventional Systems of Representation: Type II

In systems of Type II, natural signs take the place of symbols as the representational elements. A sign is given the job of doing what it (suitably deployed) can already do.

It should be remembered that what a system *represents* is *not* what its (expressive) elements indicate or mean. It is what these elements have the *function* of indicating or meaning. It is important to keep this point in mind, since the natural signs used in systems of Type II typically indicate a great many things. Normally, though, they are used to represent only *one* of these conditions—a condition which we, for whatever reason, take a special interest in and give the function of indicating. If a full tank of gas means (because of the weight of the gas) that there is a large downward force on the bolts holding the tank to the car's frame, then the fuel gauge indicates a large downward force on these bolts whenever it indicates a full tank of gas. In addition, electrically operated fuel gauges indicate not only the amount of fuel left in the tank but also the amount of electrical current flowing in the wires connecting the gauge to the tank, the amount of torque on the armature to which the pointer is affixed, and the magnitude of the magnetic field surrounding this armature. Given the way these gauges operate, they cannot indicate (i.e., have their behavior depend on) the amount of fuel in the tank without indicating (exhibiting at least the same degree of dependency on) these related conditions.

Nevertheless, we take one of these indicated conditions to be what the gauge *represents*, one of these correlated conditions to define what *kind* of gauge it is. It is, or so we say, a *fuel* gauge, not a galvanometer recording potential differences between points in the automobile's electrical wiring (though that, in a sense, is precisely what it is). Since we are interested in the amount of gasoline in the tank, not (except derivatively) in these correlated conditions, we *assign* the gauge the function of indicating the amount of gasoline in the tank. We *give* it the job of delivering *this* piece of information, calibrate and label it accordingly, and ignore the collateral pieces of information it necessarily supplies in the process. Since what an instrument or gauge represents is what it is *supposed* to indicate, what it has the *function* of indicating, and since *we* determine these functions, *we* determine what the gauge represents. If, by jacking up the fuel tank, I remove

the force on the bolts securing the tank to the car frame, the fuel gauge, though still indicating the amount of fuel in the tank, no longer indicates the amount of force on these bolts. But, under these unusual conditions, the gauge does not *misrepresent* the force on these bolts the way it could, and the way gauges sometimes *do*, misrepresent the amount of fuel in the tank. The reason it doesn't is because the gauge, even when things are operating normally, does not *represent* (though it does *indicate*) the magnitude of this force. Its *representational* efforts—and therefore its representational failures, its *mis*representations—are limited to what it has the *function* of indicating. And since the gauge does not have the function of indicating the force on these bolts, it does not misrepresent this force when it fails to indicate it. Though it is hard to imagine why we would do this, we could *give* the gauge this function. Were we to do so, then, under the unusual conditions described above, when we removed the force on these bolts by jacking up the tank, the gauge would misrepresent the force on the bolts.

It is for this reason that what the gauge represents is *partly* conventional, *partly* a matter of what we say it represents. In contrast with the case of Type I systems, however, this dependence on us, our interests and purposes, is only partial. The reason it is only partial is because the indicator functions assigned an instrument are limited to what the instrument *can* indicate, to what its various states and conditions depend on. You can't assign a rectal thermometer the job of indicating the Dow-Jones Industrial Average.[4] The height of the mercury doesn't depend on these economic conditions. The mercury and the market fluctuate independently. Trying to use a thermometer in this way is like assigning a rock the job of washing dishes.[5] My son can be given this job (even if he never does it) because he, unlike the rock, *can* wash dishes. The functions we assign to instruments are similarly restricted to what the instruments *can* do, or, if Wright (1973) is correct, what (in the case of artifacts) we *think* they can do. This makes the functions of systems of Type II restricted in a way that those of Type I systems are not restricted. It is this fact, together with the fact that once a

4. Not, at least, as an RS of Type II. One could, however, use it as an RS of Type I. Just as I used coins and popcorn to represent basketball players, and the positions and movements of these elements the position and movements of the players, there is nothing preventing one from *using* a rectal thermometer in a similar fashion to represent the Dow-Jones average.

5. For those who want to quarrel about this issue, I could, I suppose, assign a rock the job of doing my dishes if I mistook it for my son, just as I could assign a thermometer the job of indicating fluctuations in the stock market if I mistook it for something else. I do not, however, think a rock could actually *have* this function. Nor do I think a simple instrument could *have* the function of indicating something it could not indicate. This is not to say that the thermometer could not be incorporated into a more complex system that *could* indicate, and therefore could have the function of indicating, something about the stock market. But, by the same token, I could also make the rock part of a machine (pulleys, etc.) that *could* do (and, therefore, could have the function of doing) my dishes.

device has been given such a functon it performs without any help from us, that makes such systems only *partly* conventional.

The conventional, interest-relative, and purpose-dependent character of systems of Type II is especially obvious when our interests and purposes change. An altimeter represents altitude until we remove it from the aircraft for testing on the ground. It then "becomes" an aneroid barometer, representing not altitude but air pressure—something it *always* indicated, of course, but something in which we weren't interested (except insofar as it depended on, and hence served as an accurate indicator of, altitude) when flying the plane. Calibration is a process in which one's interests and purposes undergo a temporary change. *Now*, during calibration, one uses the needle's position as an indicator, not of the quantity the instrument is usually used to measure, but of the instrument's own internal condition—whether, for example, its batteries are getting weak, or whether it needs adjustment, repair, or alignment. With RSs of Type II we can, and sometimes do, change the magnitude being represented (not merely the scale for measuring a given magnitude) merely by consulting a different set of numbers on the face of the instrument. A change in the way we *use* the instrument is enough to change its function and, hence, what it represents.

One way of thinking about the difference between Type I and Type II representational systems is that in systems of Type I the function, as it were, comes first. The representational elements are given a function and then, if things go right, are *used* in conformity with this function—*used* to indicate what, relative to this function, they are supposed to indicate. I first give the dime, *its* position and movements, the function of indicating the position and movements of Oscar Robertson. Then I manipulate the dime in accordance with this assigned function. I, in virtue of my knowledge and manipulative skills, see to it that it indicates what I have assigned it the function of indicating. Not only the coin's *job* but also its *performance* of that job derives, therefore, wholly from me, the creator and user of the representational system. RSs of Type I are, then, *manifestations* or *displays* of the representational talents of their users in much the same way that a TV monitor is a *display* of the information-processing capabilities of the machinery lying behind it. With systems of Type II, however, things are different. The power of their elements to indicate comes first; their function comes second. They acquire or are assigned the function of doing one of the things they are already doing or, if not *already* doing, already *capable* of doing once harnessed in the right way. Their ability to perform their function does *not*, as in the case of systems of Type I, depend on us, on a user-system already in possession of the required indicator skills. The status of these elements as indicators is therefore *intrinsic*. What is extrinsic, and therefore still conventional, still relative to the interests and purposes of its

users, is the determination of which among the various things they can already do it is their function to do.

3.4 *Natural Systems of Representation*

A natural system of representation is not only one in which the elements, like the elements of Type II systems, have a power to indicate that is independent of the interests, purposes, and capacities of any other system, but also one in which, in contrast with systems of Type II, the functions determining what these signs *represent* are also independent of such extrinsic factors. Natural systems of representation, systems of Type III, are ones which have *their own* intrinsic indicator functions, functions that derive from the way the indicators are developed and used *by the system of which they are a part.* In contrast with systems of Type I and II, these functions are not assigned. They do not depend on the way *others* may use or regard the indicator elements.

Whatever one might think about the possibility of intrinsic functions, the type of functions that define Type III systems (a contentious point to which I will return in a moment), it is clear that what I have been calling natural signs—events, conditions, and structures that somehow indicate how things stand elsewhere in the world—are essential to every animal's biological heritage. Without such internal indicators, an organism has no way to negotiate its way through its environment, no way to avoid predators, find food, locate mates, and do the things it has to do to survive and propagate. This, indeed, is what sense perception is all about. An animal's senses (at least the so-called exteroceptors) are merely the diverse ways nature has devised for making what happens inside an animal depend, in some indicator-relevant way, on what happens outside. If the firing of a particular neuron in a female cricket's brain did not indicate the distinctive chirp of a conspecific male, there would be nothing to guide the female in its efforts to find a mate (Huber and Thorson 1985). The *place, misplace,* and *displace* neural units in the rat's brain (O'Keefe 1976), units that guide the animal in its movements through its environment, are merely internal indicators of place, of alterations in place, and of movement through a place. Such is the stuff of which cognitive maps are made, part of the normal endowment for even such lowly organisms as ants and wasps (Gallistel 1980).

The firing of neural cells in the visual cortex, by indicating the presence and orientation of a certain energy gradient on the surface of the photoreceptors, indicates the whereabouts and the orientation of "edges" in the optical input and therefore indicates something about the surfaces in the environment from which light is being reflected. The activity of these cells, not to mention comparable activity by other cells in a wide variety of

sensory systems, is as much a natural sign or indicator as are the more familiar events we commonly think of as signs—the autumnal change in maple leaves, growth rings in a tree, and tracks in the snow.

We are accustomed to hearing about biological functions for various bodily organs. The heart, the kidneys, and the pituitary gland, we are told, have functions—things they are, in this sense, *supposed to do*. The fact that these organs are supposed to do these things, the fact that they have these functions, is quite independent of what *we* think they are supposed to do. Biologists *discovered* these functions, they didn't invent or assign them. We cannot, by agreeing among ourselves, *change* the functions of these organs in the way that I can change, merely by making an appropriate announcement, what the coins and the popcorn in my basketball game stand for. The same seems true for sensory systems, those organs by means of which highly sensitive and continuous dependencies are maintained between external, public events and internal, neural processes. Can there be a serious question about whether, in the same sense in which it is the heart's function to pump the blood, it is, say, the task or function of the noctuid moth's auditory system to detect the whereabouts and movements of its arch-enemy, the bat?

Some marine bacteria have internal magnets, magnetosomes, that function like compass needles, aligning themselves (and, as a result, the bacterium) parallel to the Earth's magnetic field (Blakemore and Frankel 1981). Since the magnetic lines incline downward (toward geomagnetic north) in the northern hemisphere, bacteria in the northern hemisphere, oriented by their internal magnetosomes, propel themselves toward geomagnetic north. Since these organisms are capable of living only in the absence of oxygen, and since movement toward geomagnetic north will take northern bacteria away from the oxygen-rich and therefore toxic surface water and toward the comparatively oxygen-free sediment at the bottom, it is not unreasonable to speculate, as Blakemore and Frankel do, that *the function* of this primitive sensory system is to indicate the whereabouts of benign (i.e., anaerobic) environments.[6]

Philosophers may disagree about how best to analyze the attribution of function to the organs, processes, and behaviors of animals and plants (see, for example, Nagel 1961, Wright 1973; Boorse 1976, and Cummins 1975, all conveniently collected in Sober 1984b), but that some of these things

6. There may be some disagreement about how best to describe the function of this primitive sensory system. Does it have the function of indicating the location, direction, or whereabouts of anaerobic conditions? Or does it, perhaps, have the function of indicating the Earth's magnetic polarity (which in turn indicates the direction of anaerobic conditions)? In Dretske 1986 I described this as an "indeterminacy" of function. As long as this indeterminacy exists, there is, of course, an associated indeterminacy in what the system represents. I return to this point later.

have functions—functions, like those of the bacterium's magnetic sense or the moth's auditory sense, to be *discovered* (not invented or assigned)—seems evident not only from a common-sense standpoint but also from the practice, if not the explicit avowals, of biologists and botanists.

This is, nevertheless, a controversial topic, at least among philosophers (see, e.g., Dennett 1987), and I do not wish to rest a case for a *philosophical* thesis on what seems evident to common sense or what is taken for granted by biologists. So for the moment I take the biological examples as more or less (depending on your point of view) plausible illustrations of intrinsic functions—plausible examples, therefore, of sensory systems that, by having such functions, qualify as *natural* systems of representation. As we shall see later (chapter 4), the case for representational systems of Type III will rest on quite different sorts of functions: those that are derived, not from the evolution of the species, but from the development of the individual. Nevertheless, it is useful to think, if only for illustrative purposes, about the way certain indicator systems developed, in the evolutionary history of a species, to serve the biological needs of its members. It should be understood, though, that my use of such examples is merely an expository convenience. The *argument* that there are functions of the kind required for Type III systems, hence an argument for the *existence* of Type III systems, systems with a natural power of representation, remains to be made.

3.5 *Intentionality: Misrepresentation*[7]

Philosophers have long regarded intentionality as a mark of the mental. One important dimension of intentionality is the capacity to misrepresent, the power (in the case of the so-called propositional attitudes) to *say* or *mean* that *P* when *P* is not the case. The purpose of this section is to describe how systems of representation, as these have now been characterized, possess this capacity and, hence, exhibit some marks of the mental. Two other important dimensions of intentionality will be discussed in the following section.

Before we begin, it is perhaps worth noting that, since systems of Types I and II derive their representational powers, including their power to misrepresent, from systems (typically humans) that already have the full range of intentional states and attitudes (knowledge, purpose, desire, etc.), *their* display of intentional characteristics is not surprising. As we shall see, the traces of intentionality exhibited by such systems are merely *reflections* of the minds, *our* minds, that assign them the properties, in particular the

7. The material in this section is based on Dretske 1986. That work, and in fact this entire chapter, was heavily influenced by the important work of Stampe (1975, 1977), Millikan (1984, 1986), Enc (1979, 1982), and Fodor (1984, 1987a). Also see Papineau (1984).

functions, from which they derive their status as representations. This is not so, however, for systems of Type III. If there are such systems, *their* intentionality will not be a thing of *our* making. They will have what Haugeland (1981b) calls *original* intentionality and Searle (1980) calls *intrinsic* intentionality.

The first aspect of intentionality to be described is the capacity some systems have to represent something as being so when it is not so—the power of *misrepresentation*. It may be thought odd to accent the negative in this way, odd to focus on a system's ability to get things wrong—on its vices, as it were, instead of its virtues. There is, though, nothing backward about this approach. The ability to correctly represent how things stand elsewhere in the world *is* the ability of primary value, of course, but this value adheres to representations only insofar as the representation in question is the sort of thing that *can* get things wrong. In the game of representation, the game of "saying" how things stand elsewhere in the world, telling the truth isn't a virtue if you *cannot* lie. I have already said that indication, as I am using this word, and as Grice used the idea of natural meaning, describes a relation that cannot fail to hold between an indicator and what it indicates. There can be no *mis*indication. If the gas tank is empty, the gauge *cannot*, in this sense of the word, indicate that it is full. This is not to say that someone might not *take* the gauge as indicating a full tank. It is only to say that the gauge does not, in fact, indicate a full tank. Since indicators cannot, in this sense, fail to indicate, they do not possess the capacity of interest: the power to get things wrong. *They* don't get things wrong. *We* get things wrong by (sometimes) misreading the signs, by *taking* them to indicate something they don't. What we are after is the power of a system to say, mean, or represent (or, indeed, *take*) things as *P whether or not P is the case*. That is the power of words, of beliefs, of thought—the power that *minds* have—and that, therefore, is the power we are seeking in representational systems. Whatever *word* we use to describe the relation of interest (representation? meaning?), it is the power to misrepresent, the capacity to get things wrong, to say things that are not true, that helps *define* the relation of interest. *That* is why it is important to stress a system's capacity for misrepresentation. For only if a system has this capacity does it have, in its power to get things right, something approximating *meaning*. That is why the capacity to misrepresent is an important aspect of intentionality and why it figures so large in the philosophy of mind and the philosophy of language.

For this reason it is important to remember that not every indicator, not even those that occur *in* plants and animals, is a representation. It is essential that it be the indicator's *function*—natural (for systems of Type III) or otherwise (for systems of Type II)—to indicate what it indicates. The width of growth rings in trees growing in semi-arid regions is a sensitive

rain gauge, an accurate indication of the amount of rainfall in the year corresponding to the ring. This does not mean, however, that these rings *represent* the amount of rainfall in each year. For that to be the case, it would be necessary that it be the function of these rings to indicate, by their width, the amount of rain in the year corresponding to each ring.[8] This, to say the least, is implausible—unless, of course, we start thinking of the rings as an RS of Type II. We, or botanists, might *use* these rings to learn about past climatic conditions. Should this happen in some regular, systematic way, the rings might take on some of the properties of an instrument or gauge (for the people who use them this way). Insofar as these rings start *functioning* in the information-gathering activities of botanists as a sign of past rainfall, they may, over time, and in the botanical community, acquire an indicator function and thereby assume a genuine representational (of Type II) status. At least they might do so *for* the botanists who use them this way. But this is clearly not an RS of Type III. Though there is something in the tree, the width of the fourteenth ring, that indicates the amount of rainfall fourteen years ago, it is implausible to suppose it is the ring's function to indicate this. The variable width of the rings is merely the effect of variable rainfall. The distension of an animal's stomach is, likewise, an indicator of the amount of food the animal has eaten and (for this reason, perhaps) an indicator of the amount of food available in its environment. But this is surely not the function of a distended stomach.

This point is important if we are to understand the way RSs manage to misrepresent things. The capacity for misreprentation is easy enough to understand in systems of Type I. For here the power of the elements to misrepresent depends on *our* willingness and skill in manipulating them in accordance with the (indicator) functions we have assigned them. Since I am responsible for what the coins and the popcorn in my basketball game stand for, since I assigned them their indicator function, and since I am responsible for manipulating them in accordance with this function, the arrangement of coins and popcorn can be made to misrepresent whatever *I*, deliberately or out of ignorance, make them misrepresent. Their misrepresentations are really *my* misrepresentations.

Misrepresentation in systems of Type II is not quite so simple an affair, but, once again, its occurrence ultimately traces to whoever or whatever assigns the functions that determine the system's representational efforts. Since there is no such thing as a *mis*indication, no such thing as a natural sign's meaning that something is so when it is not so, the only way a system of natural signs can misrepresent anything is if the signs that serve as its representational elements fail to indicate something they are *supposed*

8. Fodor (1984) makes this point against Stampe's (1977) idea that the rings in a tree *represent*, in the relevant sense, the tree's age. See Stampe 1986 for a reply.

to indicate. And what they are *supposed* to indicate is what *we*, for purposes of our own, and independent of a sign's success in carrying out its mission on particular occassions, *regard* them as having (or give them) the job of doing. Without *us* there are no standards for measuring failure, nothing the system fails to do that it is supposed to do. Although the actual failures aren't *our* failures, the standards (functions) that make them failures are our standards. Putting chilled alcohol in a glass cylinder doesn't generate a misrepresentation unless somebody calibrates the glass, hangs it on the wall, and calls it a thermometer.

Only when we reach RSs of Type III—only when the functions defining what a system is supposed to indicate are intrinsic functions—do we find a *source*, not merely a reflection, of intentionality. Only here do we have systems sufficiently self-contained in their representational efforts to serve, in this one respect at least, as models of thought, belief, and judgment.

A system could have acquired the *function* of indicating that something was *F* without, in the present circumstances, or any longer, or perhaps *ever*, being able to indicate that something is *F*. This is obvious in the case of a Type II RS, where, by careless assembly, a device can fail to do what it was designed to do. As we all know, some shiny new appliances don't work the way they are supposed to work. They *never* do what it is their function to do. When what they are supposed to do is indicate, such devices are doomed to a life of misrepresentation. Others leave the factory in good condition but later wear out and no longer retain the power to indicate what it is their function to indicate. Still others, thought they don't wear out, are used in circumstances that curtail their ability to indicate what they were designed to indicate. A compass is no good in a mineshaft, and a thermometer isn't much good in the sun. In order to do what they are supposed to do, care has to be taken that such instruments are used when and where they can do their job.

The same is true of RSs of Type III. Suppose a primitive sensory ability evolves in a species because of what it is capable of telling its possessors about some critical environmental condition *F*. Let us assume, for the sake of the example, that the manner in which this indicator developed, the way it was (because of its critical role in delivering needed information) favored by the forces of selection, allows us to say that this indicator has the function of indicating *F*. Through some reproductive accident, an individual member of this species (call him Inverto) inherits his *F*-detector in defective (let us suppose inverted) condition. Poor Inverto has an RS that always misrepresents his surroundings: it represents things as being *F* when they are not, and vice versa.[9] Unless he is fortunate enough to be preserved in

9. An artificial approximation of this situation occurred when R. W. Sperry (1956) and his associates rotated, by surgical means, the eyeball of a newt by 180°. The vision of the

some artificial way—unless, that is, he is removed from a habitat in which the detection of *F*s is critical—Inverto will not long survive. He emerged defective from the factory and will soon be discarded. On the other hand, his cousins, though emerging from the factory in good condition, may simply wear out. As old age approaches, their RSs deteriorate, progressively losing their ability to indicate when and where there is an *F*. They retain their function, of course, but they lose the capacity to perform that function. Misrepresentation becomes more and more frequent until, inevitably, they share Inverto's fate.

And, finally, we have the analogue, in a Type III system, of an instrument used in disabling circumstances—the compass in a mineshaft, for instance. Consider a sensitive biological detector that, upon removal from the habitat in which it developed, flourished, and faithfully serviced its possessor's biological needs, is put into circumstances in which it is no longer capable of indicating what it is supposed to indicate. We earlier considered bacteria that relied on internal detectors (magnetosomes) of magnetic north in order to reach oxygen-free environments. Put a northern bacterium into the southern hemisphere and it will quickly destroy itself by swimming in the wrong direction. If we suppose (we needn't; see footnote 6) that it is the function of these internal detectors to indicate the whereabouts of anaerobic conditions, then misrepresentation occurs—in this case with fatal consequences.

Put a frog in a laboratory where carefully produced shadows simulate edible bugs. In these unnatural circumstances the frog's neural detectors—those that have, for good reason, been called "bug detectors"—will no longer indicate the presence or the location of bugs. They will no longer indicate this (even when they are, by chance, caused to fire by real edible bugs) because their activity no longer *depends* in the requisite way on the presence of edible bugs. Taking a frog into the laboratory is like taking a compass down a mineshaft: things no longer work the way they are

---

animal was permanently reversed. As Sperry describes it: "When a piece of bait was held above the newt's head it would begin digging into the pebbles and sand on the bottom of the aquarium. When the lure was presented in front of its head, it would turn around and start searching in the rear."

It should be noted that one doesn't disable an indicator *merely* by reversing the code—letting *b* (formerly indicating *B*) indicate *A* and *a* (formerly indicating *A*) indicate *B*. As long as this reversal is systematic, the change is merely a change in the way information is being coded, not a change in the information being coded. But though *A* and *B* are still being indicated (by *b* and *a* respectively), they are, after the inversion, no longer being accurately *represented* unless there is a corresponding change (inversion) in the way the representational elements (*a* and *b*) function in the rest of the system. This is what did not happen with the newt. It still got the information it needed, but as a result of the coding change it misrepresented the conditions in its environment.

supposed to work. Indicators stop indicating. If we suppose, then, that it is the function of the frog's neural detectors to indicate the presence of edible bugs, then, in the laboratory, shadows are misrepresented *as* edible bugs. The frog has an analogue of a false belief.[10] Occasionally, when an edible bug flies by, the frog will correctly represent it as an edible bug, but this is dumb luck. The frog has the analogue of a true belief, a *correct* representation, but no *knowledge*, no *reliable* representation. Taking a compass down a mineshaft will not change what it "says" (namely, that whichever way the needle points is geomagnetic north), but it will change the reliability, and (often enough) the truth, of what it says. Likewise, taking a frog into the laboratory will not change what it "thinks," but it will change the number of times it *truly* thinks what it thinks.

All this is conditional on assumptions about what it is the *function* of an indicator to indicate. Upon realizing that a typical fuel gauge in an automobile cannot distinguish between gasoline and water in the tank, one could insist that it is the gauge's function to register not how much gasoline is left in the tank but how much *liquid* is left in the tank. It is our job, the job of those who use the gauge, to see to it that the liquid is gasoline. If this is indeed how the function of the gauge is understood, then, of course, the gauge does *not* misrepresent anything when there is water in the tank. it correctly represents the tank as half full of liquid. And a similar possibility exists for the frog. If the function of the neural detectors on which the frog depends to find food is merely that of informing the frog of the whereabouts of small moving dark spots, then the frog is *not* misrepresenting its surroundings when, in the laboratory, it starves to death while flicking at shadows. For the internal representation triggering this response is perfectly accurate. It indicates what it is supposed to indicate: the presence and whereabouts of small, moving dark spots. The shadows *are* small moving dark spots, so nothing is being misrepresented.

Misrepresentation depends on two things: the *condition* of the world being represented and the *way* that world is represented. The latter, as we have seen, is determined, not by what a system indicates about the world, but by what it has the function of indicating about the world. And as long as there remains this indeterminacy of function, there is no clear sense in which misrepresentation occurs. Without a determinate function, one can, as it were, always exonerate an RS of error, and thus eliminate the occurrence of misrepresentation, by changing what it is *supposed* to be indicating, by changing what it is its *function* to indicate. It is this indeterminacy that

10. But not a real false belief, because, as we shall see in the next chapter, beliefs are *more* than internal representations. They are internal representations that help explain the behavior of the system of which they are a part.

Dennett (1987) dramatizes in his arguments against the idea of *original* or *intrinsic* intentionality.

What this shows is that the occurrence of misrepresentation depends on there being some principled, nonarbitrary way of saying what the indicator function of a system is. In systems of Types I and II there is no special problem because *we* are the source of the functions. We can, collectively as it were, eliminate this indeterminacy of function by agreeing among ourselves or by taking the designer's and the manufacturer's word as to what the device is supposed to do. If a watch is really a calendar watch, as advertised, then it is *supposed* to indicate the date. It "says" today is the fourth day of the month. It isn't. So it is misrepresenting the date. Case closed.

The case is not so easily closed in systems of Type III. It can only be successfully closed when internal indicators are harnessed to a control mechanism. Only by *using* an indicator in the production of movements whose successful outcome depends on *what is being indicated* can this functional indeterminacy be overcome, or so I shall argue in chapter 4.

3.6 Intentionality: Reference and Sense

If an RS has the function of indicating that *s* is *F*, then I shall refer to the proposition expressed by the sentence "*s* is *F*" as the *content* of the representation. There are always two questions that one can ask about representational contents. One can ask, first, about its reference—the object, person, or condition the representation is a representation *of*. Second, one can ask about the way what is represented is represented. What does the representation say or indicate (or, when failure occurs, what is it *supposed* to say or indicate) about what it represents? The second question is a question about what I shall call the sense or meaning of the representational content. Every representational content has both a sense and a reference, or, as I shall sometimes put it, a topic and a comment—what it says (the comment) and what it says it about (the topic). These two aspects of representational systems capture two additional strands of intentionality: the *aboutness* or *reference* of an intentional state and (when the intentional state has a propositional content) the *intensionality* spelled with an "s") of sentential expressions of that content.

Nelson Goodman (1976) distinguished between pictures *of* black horses and what he called black-horse pictures. This is basically my distinction between topic and comment. Black-horse pictures represent the black horses they are picutres of *as* black horses. Imagine a black horse photographed at a great distance in bad light with the camera slightly out of focus. The horse appears as a blurry spot in the distance. This *is* a picture of a black horse, but not what Goodman calls a black-horse picture. When

invited to see pictures of your friend's black horse, you expect to see, not only pictures of a black horse, but black-horse pictures—pictures in which the denotation, topic, or reference of the picture is *identifiably* a black horse—or, if not a *black* horse, then at least a horse or an animal of some sort.

Not all representations are pictorial. Many representations are not expected, even under optimal conditions, to *resemble* the objects they represent. Language is a case in point, but even in the case of Type II RSs it is clear that ringing doorbells do not resemble depressed doorbuttons (or people at the door) and that fuel gauges (at least the old-fashioned kind) do not resemble tanks full of gasoline. And if, as seems likely, there is in a wolf's skull some neural representation of the wounded caribou it so relentlessly follows (ignoring the hundreds of healthy animals nearby), this representation of the caribou's condition, position, and movements does not actually resemble, in the way a photograph or a documentary film might resemble, a terrified caribou. A picture, though, is only one kind of representation, a representation in which information about the referent is carried by means of elements that visually resemble the items they represent. A nonpictorial representation, however, exhibits the same dimensions. It has a reference and a meaning, a topic and a comment. My fuel gauge is not only a representation *of* an empty gasoline tank; it is also (when things are working right) an empty-tank representation. That the tank is empty is what it indicates, the information it carries, the comment it makes, about that topic. My gas tank is also very rusty, but the gauge does not comment on this feature of its topic.

The wolf's internal representation of a sick caribou may or may not be a sick-and-fleeing-caribou representation, but it certainly is a representation *of* a sick, fleeing caribou. *How* the neural machinery represents *what* it represents is, to some degree, a matter of speculation, a matter of divining what the patterns of neural activity in the wolf's brain indicate about the caribou and (since we are talking about *representations*) what, if anything, it is the function of these sensory-cognitive elements to indicate about prey. Does the wolf really represent caribou *as* caribou? Sick and lame caribou *as* sick and lame? If it turns out (it doesn't) that the wolf cannot distinguish a caribou from a moose, the answer to the first question is surely No. Perhaps the wolf merely represents caribou as large animals of some sort. Or merely as food. But the point is that unless the wolf has some means of representing comparatively defenseless caribou—a way of commenting on these creatures that is, for practical wolfish purposes, extensionally equivalent to *being a (comparatively) defenseless caribou*—its relentless and unerring pursuit of comparatively defenseless caribou is an absolute mystery, like the flawless performance of an automatic door opener that has nothing in it to signal (indicate) the approach of a person or an object. There has to be something in there that "tells" the door opener what it needs to know in

order for it to do what it does—to open the door *when* someone approaches. The same is true of the wolf.

Our ordinary descriptions of what animals (including people) see, hear, smell, feel, know, believe, recognize, and remember reflect the distinction between a representation's topic and its comment. This, I think, lends support to the idea that a cognitive system *is* a representational system of some kind, presumably a system of Type III. We say, for example, that Clyde can see a black horse in the distance without (for various reasons having to do either with the great distance, the camouflage, the lighting, or the fact that Clyde forgot his glasses) its *looking like* a black horse to Clyde, without its presenting (as some philosophers like to put it) a *black-horse appearance*. Clyde doesn't know what it is, but he thinks it might be the brown cow he has been looking for. In talking this way, and it is a common way of talking, we describe what Clyde's representation is a representation *of* (a black horse) and say how he represents it (as a brown cow). In Goodman's language, Clyde has a brown-cow representation of a black horse. At other times perhaps all we can say about how Clyde represents the black horse is as *something* in the distance. This may be the only comment Clyde's representational system is making about that topic. This isn't much different from a cheap scale's representing a 3.17-pound roast as weighing somewhere between 3 and 4 pounds. It is a rough comment on a perfectly determinate topic.

Compare Clyde's perceptual relationship to the black horse with a fuel gauge's relation to a full tank of gasoline. When things are working properly, the gauge carries information about the tank: the information that it is full. Since it is the gauge's assigned function to deliver this information, it represents the tank as full. It does not, however, carry information about *which* tank is full. Normally, of course, an automobile comes equipped with only one gasoline tank. The gauge is connected to *it*. There is no reason to comment on which topic (which tank) the gauge is making a remark about, since there is only one topic on which to comment and everybody knows this. Suppose, however, there were several auxiliary tanks, with some mechanism letting the gauge systematically access different tanks. Or suppose we were to connect (by radio control, say) Clyde's gauge to *my* tank. In this case the representation would have a different referent, a different topic, but the *same* comment. The gauge would "say" not that Clyde's tank was full but that *my* tank was full. The fact that it was saying this, rather than something else, would not be evident from the representation itself, of course. But neither is it evident from Clyde's representation of the black horse that it is, indeed, a representation of a black horse. To know this one needs to know, just as in the case of the gauge, to what Clyde is connected in the appropriate way. Examining the representation itself won't tell you what condition in the world satisfies it, what condition would (were it to

obtain) make the representation an accurate representation. For this one has to look at the wiring. In Clyde's case, there being no wires connecting him to the black horse, you have to look at the connections that *do* establish which topic his representation is a representation of. In the case of vision, that connection is pretty clearly, in most normal cases, whatever it is *from which* the light (entering Clyde's eyes) is reflected.[11]

The job of gauges and instruments is to carry information about the items (tanks, circuits, shafts, etc.) to which they are connected, not information about which item it is to which they are connected. So it is with pictures and most other forms of representation. Perceptual beliefs of a certain sort—what philosophers call *de re* beliefs (e.g., *that* is moving)—are often as silent as gauges about what it is they represent, about what topic it is on which they comment, about their *reference*. Clyde can see a black horse in the distance, thereby getting information about a black horse (say, that it is near a barn), without getting the information that it is a black horse—without, in other words, seeing *what* it is. Just as a gauge represents the gas level in my tank without representing it as the amount of gas in *my* tank, Clyde can have a belief about (a representation *of*) my horse without believing that it is (without representing it *as*) my (or even *a*) horse.

A great many representational contents are of this *de re* variety. There is a representation *of* the tank as being half full, *of* an animal as being lame or sick, *of* a doorbutton as being depressed, *of* a cat as being up a tree (or *of* a cat and *of* a tree as the one being up the other). These are called *de re* contents because the things (*re*) about which a comment is made is determined by nonrepresentational means, by means other than *how* that item is represented. That this is a picture, a photographic representation, *of* Sue Ellen, *not* her twin sister Ellen Sue, is not evident—indeed (given that they are identical twins) not discoverable—from the representation itself, from the *way* she is represented. One has to know who was standing in front of the camera to know who it is a picture of, and this fact cannot be learned (given the twin sister) from the picture itself. If causal theories are right (see, e.g., Stampe 1977), the reference of such representations will be determined by causal relations: that object, condition, or situation which is, as Sue Ellen was, causally responsible for the properties possessed by the representation (e.g., the color and distribution of pigment on the photographic paper).

Though most representations of Type II have a *de re* character, there are ready examples of comparatively simple systems having a *de dicto* content, a content whose reference is determined by *how* it is represented. Imagine a

11. Here I suppress difficult problems in the philosophy of perception, problems about the correct analysis of the perceptual object. Any responsible discussion of these topics would take me too far afield.

detector whose function it is to keep track of things as they pass it on an assembly line and to record each thing's color and ordinal position. At the time it is registering the color (red) and the position (fourth) of *delta*, it can be said that this mechanism provides a *de re* representation *of delta* as red and as the fourth item to pass by. The reference is *delta* because that is the item on the assembly line that the detector is currently monitoring (to which it is causally connected), and the meaning or sense is given by the expression "is red and number four" because that is what the detector indicates, and has the function of indicating, about the items it is presently scanning. At a later time, though, a time when the apparatus is no longer directly recording facts about delta, its representation of the fourth item as red changes its character. Its reference to delta, its representation *of* delta, now occurs via its description of delta as the fourth item. At this later time, *delta*'s color is relevant to the determination of the correctness of the representation *only insofar* as *delta* was the fourth item on the assembly line. If it wasn't, then even if *delta was* the item the detector registered (incorrectly) as the fourth item, *delta*'s color is irrelevant to the correctness of the representation. It is *the fourth item*, not *delta*, that has to be red in order for this (later) representation to be correct. Compare my belief, one day later, that the fourth person to enter the room was wearing a funny hat. If I retain in memory no other description capable of picking out who I believe to have been wearing a funny hat (as is the case with our imagined detector), then this later belief, unlike the original belief, is a belief about *whoever* was the fourth person to enter the room. I may never have seen, never have been causally connected to, the person who makes this belief true.

One can go further in this direction of separating the reference of a representation from the object that is causally responsible for the representation by equipping an RS with projectional resources, with some means of extrapolating or interpolating indicated patterns. Something like this would obviously be useful in a representation-driven control system that had a "need to act" in the absence of firm information. Imagine our detector, once again, given the function of simultaneously monitoring items on *several* assembly lines, recording the color and the ordinal value of each, and, on the basis of this information, making appropriate adjustments in some sorting mechanism. Think of it as an overworked device for weeding out rotten (nonred) apples. Since "attention" paid to one line requires ignoring the others, the device must "guess" about items it fails to "observe," or else a switching mechanism can be introduced that allows the detector to withdraw continuous attention from a line that exhibits a sufficiently long sequence of red apples. A "safe" line will be sampled intermittently, at a frequency of sampling determined by the line's past

safety record. The detector "keeps an eye on" the lines that have poor performance records, and "infers" that the apples on good lines are OK. If things are working reasonably well, this device produces a printed record containing representations of apples it has never inspected. This device has the function of indicating something about objects to which it is *never* causally related.

It is not hard to imagine nature providing animals with similar cognitive resources. Donald Griffin (1984), drawing on the work of J. L. Gould (1979, 1982), describes the way honeybees perform a comparable piece of extrapolation. Honeybees were offered a dish of sugar water at the entrance of their hive. The dish was then moved a short distance away, and the bees managed to find it. This was continued until, when the feeder was more than 100 or 200 meters from the hive, the bees began waiting for the dish beyond the spot where it had last been left, at what would be the next logical stopping place (20 to 30 meters from the last location). The bees, Griffin observes, "seem to have realized that this splendid new food source moves and that to find it again they should fly farther out from home" (pp. 206–207). The benefits of such extrapolative mechanisms are obvious. Aside from the search technique of the bees, an animal without beliefs (whether we call them anticipations, expectations, or fears) about *the next A* will not survive long in an environment where the next *A* can be dangerous.

Much more can, and should, be said about the reference or topic of a representation. But it is time to turn to its sense or meaning, *how* it represents what it represents, the comment it makes on that topic. All systems of representation, whatever type they happen to be, are what I shall call *property specific*. By this I mean that a system can represent something (call it *s*) as having the property *F* without representing it as having the property *G* even though everything having the first property has the second, even though every *F is G*. Even if the predicate expressions "*F*" and "*G*" are *coextensional* (correctly apply to exactly the same things), this doesn't guarantee that an RS will represent *s* as *F* just because it represents *s* as *G* (or vice versa). These extensionally equivalent expressions give expression to quite different representational contents. This is a very important fact about representational systems. It gives their content a fine-grainedness that is characteristic of intentional systems. It makes verbal expressions of their content *intensional* rather than *extensional*. It is this feature, together with the system's capacity for misrepresentation and the reference or aboutness of its elements, that many philosophers regard as the essence of the mental.

Representational contents exhibit this peculiar fine-grainedness because even when properties *F* and *G* are so intimately related that nothing can

indicate that something is *F* without indicating that it (or some related item) is *G*, it can be the device's *function* to indicate one without its being its function to indicate the other.[12] Nothing can indicate that *x* is red unless it thereby indicates that *x* is colored, but it can be a device's function to indicate the color of objects (e.g. that they are red) without its being its function to indicate that they are colored.

The specificity of functions to particular properties, even when these properties are related in ways (e.g., by logical or nomological relations) that prevent one's being indicated without the other's being indicated, is easy to illustrate with assigned functions, functions *we* give to instruments and detectors. For here the assignment of functions merely reflects *our* special interest in one property rather than the other. If we are, for whatever reason, interested in the number of angles in a polygon and not in the number of sides, then we can give a detector (or a *word*) the function of indicating the one without giving it the function of indicating the other even though the detector (or word) cannot successfully indicate that something is, say, a triangle without thereby indicating that it has three sides. We can make something into a voltmeter (something having the function of indicating voltage differences) without thereby giving it the function of indicating the amount of current flowing even if, because of constant resistance, these two quantities covary in some lawful way.

Though this phenomenon is easier to illustrate for Type I and Type II systems, it can easily occur, or can easily be imagined to occur, in systems of Type III. Dolphins, we are told, can recognize the shapes of objects placed in their pool from a distance of 50 feet. Apparently there is something in the dolphin, no doubt something involving its sensitive sonar apparatus, that indicates the *shapes* of objects in the water. But a dolphin that can infallibly identify, detect, recognize, or discriminate (use whatever cognitive verb you think appropriate here) cylinders from this distance should *not* be credited with the ability to identify, detect, recognize, or discriminate, say, *red* objects from this distance just because all (and only) the cylinders are red. If the fact that all (and only) the cylinders are red is a coincidence, of course, then something can indicate that *X* is a cylinder without indicating that *X* is red. This follows from the fact that an indicator could exhibit the requisite *dependence* on the shape of *X* without exhibiting any dependence on the color of *X*. But even if we suppose the connection between color and shape to be more intimate, we can, because of the different relevance of these properties to the well-being of an animal,

12. See Enc 1982 for further illustrations of this. Enc argues, convincingly to my mind, that we can distinguish between the representation of logically *equivalent* situations by appealing to (among other things) the functions of a system.

imagine a detector having the function of indicating the shape of things without having the function of indicating their color.[13]

3.7 Summary

The elements of a representational system, then, have a content or a meaning, a content or meaning defined by what it is their function to indicate. This meaning or content is a species of what Grice called non-natural meaning. These meanings display many of the intentional properties of genuine thought and belief. If, then, there are systems of Type III, and these are located in the heads of some animals, then there is, in the heads of some animals (1) something that is *about* various parts of this world, even those parts of the world with which the animal has never been in direct perceptual contact; (2) something capable of representing and, just as important, *misrepresenting* those parts of the world it is about; and (3) something that has, thereby, a *content* or *meaning* (not itself in the head, of course) that is individuated in something like the way we individuate thoughts and beliefs.

13. Taylor (1964, p. 150) notes that an experimenter can condition an animal to respond to red objects without conditioning it to respond to objects that differ in color from the experimenter's tie (which is green). He takes this to be a problem for how the property to which behavior is conditioned is selected. It should be clear that I think the answer to Taylor's problem lies, at least in part, in an adequate theory of representation, one that can distinguish between the representation of *X* as red and *X* as not green.