

# INFORMATION- THEORETIC SEMANTICS

---

FRED DRETSKE

INFORMATION-THEORETIC semantics is an attempt to ground meaning—as meaning is understood in the study both of language and of mind—in an objective, mind- (and language-) independent notion of information. It is typically part of a larger philosophical effort to naturalize mentality—to understand thought and its expression in language as just another strand in the fabric of material affairs.

If symbols are the bearers of meaning, the things in the world that *have* meaning, information-theoretic semantics locates the primary source of this meaning in the relations these symbols bear to the world they are, or purport to be, about. These symbol–world relations can be described in explicitly information-theoretic terms (source, signal, noise, receiver, etc.), as occurs in Dretske (1981), or they can be described in more general causal terms (Stampe 1977, 1986; Matthen 1988; Fodor 1990; Israel and Perry 1990). However it may be expressed, though, the basic idea is that the primary determinant of a symbol’s meaning is those situations in the world about which symbols carry, or are supposed to carry (more about this important qualification in a moment), information, those situations in the world of which the symbols are (in normal conditions) reliable signs. Since, according to some (e.g. Dretske 1986; Israel and Perry 1990), the information an event carries is what its occurrence indicates, informational semantics is sometimes referred to as *indicator semantics*: what a symbol means is what, in certain central cases, it indicates, or is supposed to indicate, or normally indicates, about other parts of the world. Theories of this sort are to be contrasted with *conceptual-role*, *procedural*, or

*consumer* semantics—theories that locate the meaning of a symbol not (or not *only*) in upstream causes, but also (and sometimes *primarily*) in the downstream effects these symbols have on other symbols and on the behaviour of the system in which they occur (Block 1986; Papineau 1987; Millikan 1989).

Informational semantics takes the primary—at least the original—home of meaning to be the mind: meaning as the content of thought, desire, and intention. The meaning of beliefs, desires, and intentions is what it is we believe, desire, and intend. The sounds and marks of natural language derive their meaning from the communicative intentions of the agents who deploy them. As a result, the information of chief importance to informational semantics is that occurring in the transactions between animals and their environments. So for informational semantics the very existence of thought and, thus, the possibility of language depends on the capacity of (some) living systems to transform information (normally supplied by perception) into meaningful (contentful) inner states like thought, intention, and purpose.

## 22.1 INFORMATION

---

The word 'meaning' is ambiguous. Two of its possible meanings (Grice 1989) are: (1) *non-natural* meaning—the sense in which the English word 'fire' stands for or means fire; and (2) *natural* meaning—the way in which smoke (not the word 'smoke', but smoke itself) means (indicates, is a sign of) fire. Non-natural meaning, the kind of meaning theories of meaning (like information-theoretic semantics) are supposed to be theories of, has no necessary connection with truth. The words 'Jim has the measles' mean that Jim has the measles whether or not Jim has the measles. Natural meaning, on the other hand, requires the existence of the condition meant: if Jim does not have the measles, the red spots on his face do not mean (indicate) that he has the measles. Perhaps all they mean is that he has been eating too much candy. Natural meaning, what an event indicates, what it is a sign of, is a relation between the sign and what it signifies that does not depend on anyone recognizing or identifying what is meant. It does not depend on conscious beings at all. Expanding metal (e.g. the mercury in a thermometer) indicates or means that the temperature is increasing whether or not anyone knows that this is what it means. A particular cloud formation means a cold front is approaching even if (because no one is aware of its meaning) it does not mean this *to* anyone. We found out that this is what it means. Unlike non-natural meaning, what things indicate, what they mean in a natural sense, is something we learn about events by patient investigation. We learn what tracks in the snow (or a cloud chamber) mean; we do not, as we do with words and other conventional symbols, assign them their meaning.

Information, as this is used in informational semantics, is akin to natural meaning. It is what the signal, some event occurring at a 'receiver' (an animal's brain, for instance), indicates about conditions existing at a causal 'source' (perhaps an object in the animal's environment). It might, for instance, mean (indicate, carry the

information) that the object is moving, that it is located between two other objects, or that it is changing colour. The information carried by a signal, as so understood, is a fact—the fact expressed by the true proposition describing what the signal indicates about the source. So if  $e$ , an event at the receiver, indicates that  $s$ , an object at the source, is moving to the left, this fact about  $s$  is the information  $e$  carries about  $s$ . To speak about information is to speak about facts—the true propositions about a source that we (if properly attuned) can come to know about that source by receipt of a signal carrying that information. That, indeed, is why information is such an important commodity. It is why we, human beings, pay money (for tuition, for instance) to obtain it, why (in wartime) people are tortured in order to extract it from them, and why billions are spent on scientific instruments to obtain increasingly precise and exotic forms of it. The tendency in computer science to construe information as anything—whether true or false—capable of being stored on a hard disk is to confuse non-natural meaning (or perhaps just structure) with genuine information. It leaves it a mystery why information should be thought a useful commodity. Information booths are not merely places where one goes to hear people utter meaningful sentences. They are places you go expecting people to utter meaningful *true* sentences about the topics they dispense information about.

Thinking of information in terms of natural signs and what they indicate makes it clear that information depends on a system of stable relations existing between signal and source. It isn't enough, for instance, that  $s$ , an object at a source, always happens to be moving whenever a particular type of event,  $e$ , occurs at the receiver. That isn't enough to make  $e$  indicate that  $s$  is moving. No, for  $e$  to mean (in the natural sense) that  $s$  is moving, for  $e$  to carry this piece of information,  $e$  must depend on  $s$ 's movement in a particularly reliable way. Circumstances must be such that in these circumstances *only*  $s$ 's movement will result in  $e$ . If, in these conditions, something else can result in  $e$ , then the occurrence of  $e$  does not mean that  $s$  moved. The ringing of my telephone doesn't indicate, doesn't, therefore, carry the information, that *you* are calling me simply because, in point of fact, you are the only person who ever calls me. What is relevant is not whether anyone else ever *does* call me, but whether anyone else *might* call me—even if only someone dialing a wrong number or a bothersome telemarketer. If it might be someone else, then the ringing phone does not indicate that *you* are calling me. At most it means that it is probably you.

Another way of expressing this relation between source and receiver, the relation on which the flow of information depends (one I adopted in Dretske 1981 in order to make clear its connection with the mathematical theory of information and cognitive science in general), is to say that  $e$ , some event at a receiver, carries information about an object,  $s$ , at the source—the information, say, that  $s$  is  $F$ —only if conditions are such that  $e$  raises the probability of  $s$ 's being  $F$  to 1. It isn't enough to raise the probability to something less than 1—to 0.99 for instance. That isn't good enough. The reason it isn't good enough is that the relation in question (whether called information, indication, or natural meaning) is a transitive relation: if  $a$  indicates (means, carries the information) that  $b$ , and  $b$  indicates (means, carries the information) that  $c$ , then  $a$  must indicate (mean, carry the information)

that *c*. If force on the restraining spring means that an electrical current is flowing in the circuit, and current flow in the circuit means there is a voltage difference, then force on the spring indicates a voltage difference. If Sally's expression indicates she is interested, and interest indicates (at least partial) understanding, then Sally's expression indicates (at least partial) understanding. This is why no signal can carry the information that water is freezing without carrying all the information carried by freezing water—that, for example, the temperature is at or below 32 °F. Setting the probability (required for the transmission of information) at anything less than 1 would not preserve this transitivity. It is not (in general) true that if the probability of *b*, given *a*, is (at least) 0.99, and the probability of *c*, given *b*, is also (at least) 0.99, then the probability of *c*, given *a*, is also (at least) 0.99. It may be less than 0.99. So if we are going to use probability at all to express the relations in question, those relations between source and receiver that enable events at the receiver to carry information about the source, we *must* express these as probabilities of 1. Nothing less than 1 maintains the transitivity of this indicator (natural meaning, informational) relation.

It may be thought that this sets the bar too high. In our messy unpredictable world probabilities seldom, if ever, reach a value of 1. If information requires a probability of 1, then we seldom, if ever, get information. This worry betrays a misunderstanding. The probabilities in question, whether they are probabilities of 1, 0.99, or 0.5, are assigned against a background of stable circumstances, circumstances that can, but are assumed not to, change for purposes of fixing what can happen. In normal circumstances the ring of my doorbell carries information: it indicates or means that someone is at my door. It does so despite the fact that in different conditions—when there is a short circuit in the wiring, when there is an infestation by poltergeists, when squirrels have been trained to push doorbell buttons—a doorbell can be made to ring without anyone (a person) pushing the doorbell button. But in normal circumstances, circumstances that include the actually existing wiring and the absence of poltergeists and trained squirrels, there is nothing else besides someone pressing my doorbell button that can make my doorbell ring. So in these circumstances the ring indicates that someone is at my door. It carries this information. It makes the probability of someone's being at my door 1. It makes the probability 1 whether or not we know it is 1. If it doesn't make the probability 1, if there is, unknown to us, a small probability that my doorbell is ringing when no one is at my door (there is a trained squirrel that ranges freely in the neighbourhood), then the ring does not indicate that someone is at my door. All it indicates is that someone is probably there. It might be the squirrel.

## 22.2 MEANING

---

Semantics is the study of meaning in so far as meaning is understood to be *non-natural* meaning: the kind of meaning in which 'smoke' means smoke (not fire); the

kind of meaning in which a prophet can say, and thus mean, that the end is near without the end actually being near; the kind of meaning in which something in a person's head, for instance a perceptual experience or a belief, can mean (represent) one line to be longer than another even though the lines are of equal length. Information, as we have just described it (i.e. as indication or natural meaning), obviously cannot be equated with meaning in this (non-natural) sense. Nothing can carry the information that one line is longer than another, nothing can (naturally) mean or indicate that this is so, unless, in fact, the line is longer than the other. So an information-theoretic semantics, if it is to do the job, if it is to provide a plausible theory of non-natural meaning, must supplement or combine information as this is presently being understood with some other ingredient to capture the targeted concept.

There is some disagreement about what this additional idea or ingredient might be, but, whatever it is, it must somehow manage to convert natural signs of F, things that carry information about the F-ness of things, into things, symbols, which do not (or need not) carry such information. The favoured way of doing this is by appealing to the sort of 'norms' generated by things having a certain function. The basic idea is modelled on the way we give certain objects—measuring instruments, for instance—the power to mean, in a non-natural way, that something is so-and-so by giving these objects the job or function of 'telling us' (carrying the information) that something is so-and-so. A speedometer can say or represent that a car is going 60 mph when it isn't. What gives the instrument the power to mean, non-naturally, that the car is going 60 mph is the fact that it has a certain informational function (a function we—designers, makers, and users—give it): the function of telling us, providing us with information, about the speed of the car. When it is doing its job, it tells us how fast the car is going. If, because of mechanical difficulties, the instrument ceases to do its job, it nonetheless still 'says' (represents) that the car is going a certain speed, but it now misrepresents the speed. It (by pointing at the numeral '60') says something false. It exhibits a form of non-natural meaning.

No one, however, gives our brains an information-carrying function in the way that we give measuring instruments their information-carrying functions. So if this is to be a model for how our brains acquire (in the form of experience and thought) the power to mean that something is so (whether or not it is so), the power to think that one line is longer than another when it isn't, the information-carrying functions must be (what we may call) natural functions—perhaps biological functions—functions that do not depend on or derive from us in any way. To smuggle in functions that, in whatever way, depend on our purposes in the way the functions of speedometers depend on them is to smuggle in at the very beginning of the analysis the very thing—the kind of meaning associated with human purposes and thought—that our theory of meaning is supposed to be an analysis of.

Setting aside for the moment questions about the existence of natural functions (I return to this in the next section), the idea would be to equate meaning or representational content with an element's information-carrying function. The functions convert natural signs into non-natural symbols. If event type E has tokens,

$e_1, e_2, \dots$  which, in normal circumstances, are natural signs that  $s$  is  $F$ , and if  $E$  acquires, through an appropriate history, the function of carrying this information, then tokens of that type thereafter mean, in a non-natural way, that  $s$  is  $F$  even when  $s$  isn't  $F$ . They mean (non-naturally) this because that is what they are supposed to mean (naturally). The norm invoked by speaking of this as something they are *supposed* to mean is given by their information-carrying function. They are supposed to mean this in the same way hearts are supposed to pump blood.

Just as a cashier, someone whose job is to operate the cash register, can be asked to sweep floors, a symbol for  $F$ , something whose job it is to indicate  $F$ , can be pressed into kinds of service unrelated to its information-carrying function. It might appear, for instance, in a desire for  $F$ , a fear of  $F$ , or just idle thought about  $F$ . Once we have a word for  $F$  in the language of thought, this word can now occur in mental questions (I wonder whether that is an  $F$ ), commands (Give me an  $F$ ), and hopes (I hope that is an  $F$ ), as well as assertions (judgements that something is an  $F$ ).

## 22.3 PROBLEMS AND PROMISES

---

That is the good news. Now for some bad news. How bad the news is depends on how intractable the following problems are. I make suggestions here and there, but sometimes, with the customary excuse about lack of space, I provide nothing more than a promise about how an answer might go. The ultimate plausibility of information-theoretic semantics depends on there being detailed and convincing solutions to these problems.

### 22.3.1 Natural Functions

I spoke above of natural functions, the kind of function a thing has that is independent of anyone's recognition and/or attribution of that function. Stop signs and speedometers have functions, but they are not natural functions. As we all know, what stop signs and speedometers are supposed to do, or what we are supposed to do with them, derives from the collective purposes, desires, and beliefs of their designers, makers, and users. Information-based semantics needs something else. It needs natural functions, functions that (to avoid circularity) are independent of collective (or individual) purposes and beliefs. Where are these functions? What is it *in nature* that gives the visual systems of animals the function of providing information about optical surroundings? What makes the provision of information something the eyes, ears, and nose are for? Granted they *do* supply information; we couldn't survive without it. The question is whether there is any sense in which they are *supposed* to supply it.

The presumption is, of course, that the functions have their source in natural selection and learning. If the heart is supposed to pump blood and the liver is supposed

to clean it, if these are the functions of these bodily organs created by a process of natural selection, why can't the eyes and ears and associated neural systems also have duties—in particular, information they are supposed to provide? Or if during operant conditioning, a pervasive form of learning, internal indicators of those stimulus conditions with which reinforced behaviour is to be coordinated come to exercise control over behaviour (the only way this kind of learning can be successful), why isn't it thereafter the job, the function, the purpose of these indicators (e.g. perceptual states) to provide information about the stimulus conditions on which the acquired behaviour depends?

It isn't important that the word 'function' be used to describe what these organs or systems develop to do. Some people seem to think that the word 'function' (in the sense of 'purpose', not in the sense of 'causal role') is only really appropriate when applied to systems to which human beings, given their special interests, have assigned a purpose.<sup>1</sup> According to this view, there is, independent of our interests, nothing the eyes, ears, and nose are for, nothing that they are supposed to be doing. We needn't, however, quarrel over the word 'function'. What is important for the purposes of information-theoretic semantics is that there be a set of circumstances, or perhaps a kind of history, that, independent of human interests, grounds descriptions of animals and their parts as ill, sick, broken, damaged, injured, diseased, defective, flawed, infected, contaminated, or malfunctioning. If the truth of these descriptions is independent of *our* interests and purposes,<sup>2</sup> then there is a way natural systems are supposed to be, or supposed to behave, that is independent of how we conceive them. There would, therefore, be a perfectly natural sense in which, given appropriate development, the information-processing systems in animals would be able to make mistakes—able, that is, to mean that something was so when it wasn't. Misrepresentation, the sine qua non of non-natural meaning, would thus become possible in a perfectly natural way. It is hard to see why this much isn't available.

### 22.3.2 The Disjunction Problem

In order to mean, in the non-natural way we are trying to understand, that *s* is *F*, it must be possible for something to mean this when *s* isn't *F*. It doesn't have to be a dog for me to think and say it is. Symbols (linguistic or mental) that mean dog don't have to be caused by dogs. They can be caused by wolves or (Fodor's example<sup>3</sup>) cats-on-a-dark-night—things one might mistake for a dog. But anything that means or indicates, anything that carries the information, anything that is a sign, that *s* is a dog

<sup>1</sup> For an excellent collection of essays debating this issue see Arieu et al. (2002).

<sup>2</sup> Is an animal sick or diseased, or is its leg broken, only if we, human beings, deem it so? That doesn't sound right. If these conditions really are independent of our (or, indeed, anyone's) beliefs and interests, then there is a way things are supposed to be that is, in the appropriate sense, natural or objective in the way required by information-theoretic semantics.

<sup>3</sup> It was Fodor who first raised the disjunction problem (1984). See also Fodor (1990).

has to be caused by a dog.<sup>4</sup> This is the fundamental difference between natural and non-natural meaning.

This seems to create a problem. How can you convert something that is a sign of F, something that indicates that F, something that (carrying information about F-ness) can only be caused by an F, into something, a symbol for F, that can be caused by non-Fs?

Fodor (1990: 60) was certainly right that almost all attempts to solve this problem rely on distinguishing two contexts—one (call it the pure or normal context) in which the symbol for F is only caused by Fs (thus carrying information about the F-ness of a source) and one, call it the impure context, in which it can be caused by a variety of non-Fs (those, for instance that look like Fs), thus making misrepresentation (and, therefore, non-natural meaning) possible. The symbol type gets its information-carrying function, and thus its meaning, in the first context, when it is carrying information. This gives it a meaning that it takes to the impure context, a context in which, given its acquired function, it now means F (what it has the function of indicating) without anything necessarily being F.

The problem is to understand how 'pure' cases are possible, the conditions in which something can indicate dog while, at the same time, being capable of being caused (in the impure condition) by a wolf. If in impure conditions the symbol can be caused by a wolf (or anything else one might mistake for a dog), then why wouldn't it have been caused by a wolf in the pure case if, contrary to fact, a wolf had appeared there? If it would have been caused by a wolf had a wolf appeared in the pure condition, then how, in the normal or pure case, can the symbol indicate, how can it carry the information, that something is a dog? All it really seems to indicate, given the constraints on information described earlier, is a disjunctive (hence the 'disjunction' problem) fact, the fact that *s* is either a dog or a wolf. If that disjunctive fact is, indeed, the information it really carries, then the information the symbol acquires the function of carrying must be this disjunctive information. According to information-theoretic semantics, then, this disjunctive fact is what the element comes to (non-naturally) mean. If that is what it non-naturally means, however, then instances of the symbol caused by wolves in the impure condition are not misrepresentations. The cause—a wolf—actually is what the symbol's meaning says it is; namely, a wolf or a dog. So we have not achieved what we were after; that is, a symbol, something that can mean dog when it is caused by, and thus applied to, a wolf.

A lot of ink has been spilt on this problem, and nothing I can say in the space of a few paragraphs is going to settle matters, but I think it worth emphasizing that natural meaning, indication, and, therefore, information (of the relevant sort) are all context-dependent notions. Something can indicate F in one set of circumstances,

<sup>4</sup> This isn't strictly true, since a signal can carry the information that *s* is F without being caused by *s*'s being F. There are causal arrangements (e.g. common causes) in which *e* carries the information that *s* is F without being caused by *s*'s being F. The claim in the text, though, is close enough to being true for the purposes at hand. Nothing can indicate or mean (naturally) that *s* is a dog, nothing can carry this information about *s*, unless, in fact, *s* is a dog.

not in another. Certain coloration and markings on the bird indicate that it is a blue jay, but these signs can (be made to) occur on an object that is not a blue jay. If circumstances are such that fake jays (non-blue jays that have the same colour and markings) are a genuine possibility, then coloration and markings of that sort do not indicate that the bird is a blue jay. They only indicate something disjunctive in character—that it is either a blue jay or a fake jay. Readers familiar with recent developments in epistemology will recognize these points as part of the ‘relevant-possibility’ theme in theories of knowledge. Recognizing a blue jay at the bird feeder does not require that one be able to distinguish real blue jays from fake jays—not if fake jays are not, as they usually aren’t, relevant possibilities under normal viewing conditions. That is why knowing there is water (i.e.  $H_2O$ ) in the creek does not require being able to distinguish  $H_2O$  from XYZ (a molecularly distinct substance found on Twin Earth that looks like water). Not if XYZ is not, as it clearly is not if it is found *only* on Twin Earth, a relevant possibility. If, on the other hand, XYZ is actually found in some lakes and streams on earth, then you don’t know the water you see is water even if it is water. It might, for all you can tell, be XYZ.

Quite independently of theories of meaning, then, we must recognize that something can carry information about dogs (blue jays, water) in one context, not in another. There are pure contexts, contexts in which dogs have to be taken as the only relevant cause of (the symbol) DOG, and impure contexts, contexts in which wolves, cats-on-a-dark-night, and perhaps even drugs in the bloodstream might be causing DOG. This being so, I see no problem in appealing to this distinction in a theory of meaning. A symbol acquires the function of carrying information in one context, a pure context, and it is used in contexts, impure contexts, in which it may or may not carry the information it has the function of carrying, contexts in which it can, therefore, get things wrong. My own view is that the pure contexts are those in which an information-carrying element is incorporated into a control system, a context defined by that time in which the information (relevant to an organism’s adaptive behaviour) ‘gets its hand on the steering wheel’ (see Dretske 1981, 1988). But there are other possibilities.

### 22.3.3 The Grain Problem

Not only do information-carrying functions give physical systems (and, therefore, the central nervous system) the power to misrepresent the world—thus serving as one way natural meaning is converted into non-natural meaning—they also hold promise of solving a related problem that any naturalized semantics faces: the problem of *grain*. Non-natural meaning individuates propositions in a very fine-grained way. Thus, for instance, ‘*b* is 32 °F’ means (non-naturally) something different from ‘*b* is 0 °C’, even though these are merely two different ways of picking out the same temperature: the freezing point of water. In saying that *b* is 32 °F one does not say that *b* is 0 °C. One who believes that *b* is 32 °F does not necessarily believe that *b* is 0 °C. These are two distinct beliefs, distinct meanings, distinct mental contents. Yet

if something is a sign (indicates, carries the information) that  $b$  is  $32^{\circ}\text{F}$  it must be a sign (indicate, carry the information that)  $b$  is  $0^{\circ}\text{C}$ . We may not know it carries this information, of course, but it nonetheless carries it. If the probability, given the signal, of  $b$  being  $32^{\circ}\text{F}$  is 1, then the probability that  $b$  is  $0^{\circ}\text{C}$  is also 1. Natural meaning does not carve things up as finely as does non-natural meaning. This appears to be a problem for theories like information-theoretic semantics that propose to analyse non-natural meaning in terms of natural meaning. Non-natural meaning and natural meaning have different grains.

And so they do, but information-theoretic semantics is not making the mistake of identifying non-natural meaning with natural meaning. Non-natural meaning is being identified with an object's information-carrying function, and it is reasonably clear that information-carrying functions can carve up propositional space in a more finely grained way than does information. Of the many things a heart does, only one of these, pumping blood, is its biological function. So too, perhaps, although an internal state is indifferent to how the facts it carries information about are described (that  $b$  is  $32^{\circ}\text{F}$  or  $0^{\circ}\text{C}$ ), it may have the function of carrying this information in terms of a value on a Fahrenheit rather than a centigrade scale. Enc (1982) gives the example of a mechanical device that, given the way it does its job (by counting lines, not angles), has the function of detecting trilaterals, not triangles, even though, geometrically speaking, these are the same. If it is possible for functions to pull apart, in this way, facts that are, informationally speaking, equivalent, non-natural meanings and informational functions could exhibit the same propositional grain.<sup>5</sup> There would therefore be no principled objection to equating non-natural meaning, as this appears in the form of thought and intention, with information-carrying functions even though information and meaning (of the sort being analysed) have very different grains.

## 22.4 EPIPHENOMENALISM

---

If there really are semantic engines, systems whose behaviour (some of it anyway) is driven (explained) by the semantic properties of their internal states, a theory of meaning should reveal just how meaning gets its hands on the steering wheel. How does meaning achieve this causal relevance? If the neurobiological properties of internal events are not enough to explain why we do some of the things we do—the purposeful actions—how do the semantic properties, the content or meaning of internal events (i.e. what we believe, intend, and desire), figure in causal explanations of behaviour?

<sup>5</sup> See Dretske (1981: 215–22) for a similar discussion of how to distinguish, on information-theoretic grounds and compositional considerations, the concept (and, therefore, belief that something is) WATER from the concept (belief that something is)  $\text{H}_2\text{O}$  despite the (metaphysically) necessary equivalence of water with  $\text{H}_2\text{O}$ .

The problem of finding a genuine explanatory role for meaning in the behaviour of intentional (i.e. semantic) systems is especially difficult for semantic theories that locate the source of meaning in the extrinsic or relational properties of the internal events that presumably have meaning. If a theory says that the meaning or content of internal events derives from the way these symbols are related—functionally, informationally, causally, or whatever—to (usually) external affairs, then it should be possible for physically indistinguishable events, symbols that have all their intrinsic properties in common, to have quite different meanings. This should be possible in the same way it is possible for identical twins, say, to have different histories. According to orthodox thinking about causality, however, the causally relevant properties of an object are local or intrinsic. Physically indistinguishable objects (a real \$100 bill and a perfect forgery, say), placed in the same circumstances, will have exactly the same effects on the rest of the world. They will have different histories, but these historical (extrinsic, relational) differences will be masked or screened off by their current physical constitution. And so it is with meanings in so far as meaning supervenes on history. What will be causally relevant to the effects of symbols in a system will not be their history—and, therefore, their meaning. It will be their physical (intrinsic) properties.

Information-theoretic semantics has this problem in spades, since it locates the meaning of a symbol in the history of that symbol, in what kind of information it was, at some point in the past, developed to carry, and history is obviously not a local, not an intrinsic, and, therefore, not (according to orthodox theory) a causally relevant property. Every other semantic theory (and this seems to be most theories) that locates the meaning of a symbol in its extrinsic or relational properties has the same problem. This problem is most often dramatized by Davidson's swamp-man example. Lightning strikes a swamp creating a duplicate of Donald Davidson, a being that is physically indistinguishable from Davidson but one having a completely different history. Does one's theory of meaning assign the same meanings, the same content, to the internal states of this duplicate being as it does to Donald Davidson? If so, then history is irrelevant to the assigned meanings. If not, then the duplicate has no thoughts. It is a zombie. The second choice seems to be inconsistent with scientific materialism. The first choice is inconsistent with information-theoretic semantics. Conclusion? Information-theoretic semantics is inconsistent with scientific materialism. This wouldn't be so bad for some theories, but for information-theoretic semantics it is a disaster, since it was proposed as a materialistically acceptable account of meaning.

Some philosophers (e.g. Paul Churchland 1981; Patricia Churchland 1986) take this problem to be insurmountable and construe semantic explanations of behaviour, folk-psychological explanations couched in terms of the content or meaning of internal events, as, at best, a *façon de parler*, a convenient verbal heuristic that will eventually (when we know enough) be replaced by scientific (neurobiological) explanations of behaviour. Others (e.g. Dennett 1987) acknowledge the instrumental or heuristic nature of semantic explanations, but take the heuristic—the intentional stance—to be unavoidable. Still others argue for a genuine relevance

for meaning in a variety of different ways: Fodor (1987) identifies an alleged kind of meaning, narrow meaning, that is intrinsic to the events that have it; Dretske (1988) argues that the behaviours to be explained by meaning are not the bodily movements that are best explained by neurobiology, but, rather, causal processes that result in bodily movements (which are best explained by the extrinsic properties of the internal causes); Kim (1996) identifies a kind of causality, supervenient causality, that extrinsic properties (and, therefore, meanings) can participate in; Burge (1989) defends the causal relevance of semantic properties by arguing that genuine causal explanations are those that are actually given and accepted in our daily explanatory practice (a practice loaded with semantic explanations).

Perhaps, though, the best defence against the charge of epiphenomenalism is to say that although the problem is real enough, it is not just a problem for information-theoretic semantics. It is a problem for any theory of meaning—and this seems to be *all* theories of meaning—in which meaning resides in a symbol's extrinsic or relational properties. In philosophy everybody's problem isn't anybody's problem.

## REFERENCES

- Ariew, A., Cummins, R., and Perlman, M. (2002) (eds.), *Functions* (Oxford: Oxford University Press).
- Block, N. (1986), 'Advertisement for a Semantics for Psychology', *Midwest Studies in Philosophy*, 10: 615–78.
- Burge, T. (1989), 'Individuation and Causation in Psychology', *Pacific Philosophical Quarterly*, 70: 303–22.
- Churchland, P. M. (1981), 'Eliminative Materialism and Propositional Attitudes', *Journal of Philosophy*, 78: 67–90.
- Churchland, P. S. (1986), *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (Cambridge, Mass.: MIT Press).
- Dennett, D. (1987), *The Intentional Stance* (Cambridge, Mass.: MIT Press).
- Dretske, F. (1981), *Knowledge and the Flow of Information* (Oxford: Blackwell).
- (1986), 'Misrepresentation', in R. Bogdan (ed.), *Belief: Form, Content, and Function* (Oxford: Clarendon), 17–36.
- (1988), *Explaining Behaviour*. (Cambridge, Mass.: MIT Press).
- Enc, B. (1982), 'Intentional States of Mechanical Devices', *Mind*, 91: 161–82.
- Fodor, J. (1984), 'Semantics, Wisconsin Style', *Synthese*, 59: 231–50.
- (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, Mass.: MIT Press).
- (1990), *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press).
- Grice, P. (1989), *Studies in the Way of Words* (Cambridge, Mass.: Harvard University Press).
- Israel, D., and Perry, J. (1990), 'What Is Information?', in P. P. Hanson (ed.), *Information, Language, and Cognition* (Vancouver: University of British Columbia Press), 1–19.
- Kim, J. (1996), *Philosophy of Mind* (Boulder, Col.: Westview).
- Matthen, M. (1988), 'Biological Functions and Perceptual Content', *Journal of Philosophy*, 85: 5–27.
- Millikan, R. (1984), *Language, Thought, and Other Biological Categories* (Cambridge, Mass.: MIT Press).

- 
- (1989), 'Biosemantics', *Journal of Philosophy*, 86: 281–97.
- Papineau, D. (1987), *Reality and Representation* (Oxford: Blackwell).
- Stampe, D. (1977), 'Towards a Causal Theory of Linguistic Representation', *Midwest Studies in Philosophy*, 2: 42–63.
- (1986), 'Verificationism and a Causal Account of Meaning', *Synthese*, 69: 107–37.

## CHAPTER 23

---

# BIOSEMANTICS

---

RUTH GARRETT MILLIKAN

'BIOSEMANTICS' was the title of a paper on mental representation originally printed in *the Journal of Philosophy* in 1989. It contained a much-abbreviated version of the work on mental representation in *Language, Thought, and Other Biological Categories* (Millikan 1984). There I had presented a naturalist theory of intentional signs generally, including linguistic representations, graphs, charts and diagrams, road-sign symbols, animal communications, the 'chemical signals' that regulate the functions of glands, and so forth. But the term 'biosemantics' has usually been applied only to the theory of mental representation. Let me first characterize a more general class of theories called 'teleological theories of mental content' of which biosemantics is an example. Then I will discuss the details that distinguish biosemantics from other naturalistic teleological theories.

Naturalistic theories of mental representation attempt to explain, in terms designed to fit within the natural sciences, what it is about a mental representation that makes it represent something. Frequently these theories have been classified as either picture theories, causal or covariation theories, information theories, functionalist or causal-role theories, or teleological theories, the assumption being that these various categories are side by side with one another. But they are not. Teleological theories are *specific forms* of one or another, or of some combination, of the other kinds of theories. What teleological theories have in common is not any view about the nature of representational content; that is, about what makes a mental representation represent something. What they have in common is only a view about how falseness in representations is possible.

Roughly, the idea is this. You tell the teleologist what you think makes some item in the head, some facet or activity of the brain, into a representation of some facet of the world, say, into a belief that it is raining. The teleologist may well agree with your theory about this. But then she will go on to point out (typically this is so)

that your theory is really, at root, a story only about what it is for a mental state or activity to represent *truly* or *correctly*. You need to add a story about what a representation is like that represents falsely. And she will claim that this is very easy to do. Assume that the brain was designed, by evolution or learning, to make or to learn to make representations of the kind you have described. But what it was designed to do will not always be what it in fact does. Difficult environmental circumstances, even circumstances that merely fail to be ideal, often cause temporary failures for biological systems. Systems designed to produce representations will sometimes fail to produce them correctly. Sometimes they will produce items that behave in the mind/brain as though they represented something, but that in fact do not represent anything. These are false representations. They are 'false' in the dictionary sense of 'not genuine or real', 'resembling but not accurately or properly designated as such', the sense in which false faces and false fronts are false. That is, teleological theories are best understood as denying that there IS any state of affairs or occurrence being represented when one thinks falsely or that there IS any object at all that is being represented when one thinks emptily, say, when seeming to think about 'phlogiston' or 'the ether'.

Similarly, there IS no object, not even an inner one, being seen when one has a hallucination. Mistaken representations, rather than representing peculiar objects, things called 'intentional contents', are just representations that are failing to represent. False representations are representations yet fail to represent in the same way that something can be a can opener but be too blunt, hence fail to open cans or—and this is a better analogy—something can be a coffee maker yet fail to make coffee because the right ingredients were not put in or it was not turned on. They are 'representations' only in the sense that the biological function of the cognitive systems that produced them was to make representations. Thus, falsehood is explained by the simple fact that biological purposes often go unfulfilled, and the ghostly realm of intensions, reified meanings, non-existent objects of thought, intentional objects, propositional contents, and so forth is cleanly swept away. Strictly speaking, you can't represent something that doesn't exist.

But much work still remains. The teleologist must give an account of biological functions and of functions derived from learning that can support the view that the mature brain has the production of representations, representations of quite specific kinds, as one of its functions. The account I have given ultimately rests these functions on a variety of different kinds of selection, the most fundamental being natural selection. That story can be found in Millikan (1984: chs. 1–2; 1993: chs. 1–2; 2002). In Millikan (2004: ch. 1 n. 2) it is defended against the claim that human intelligence could have been a genetic accident rather than a trait selected for.

The other main work that remains is to explain what it is for some facet or activity of the brain to be a representation of some affair in the world. What teleological theories do not have in common is any agreed-on description of what representing is. They do not agree on what an organism that is representing things correctly, actually representing things, is doing, hence on what it is that an organism that is misrepresenting is failing to do. Teleological theories, just as such, are not theories of mental

content. Failure to grasp this last point has led many to take a dismissive attitude toward teleological theories. How, they ask, could the question whether my current thought is the thought that grass is green rather than the thought that aardvarks bark be a matter that is settled in part by reference to evolutionary history or to my past learning history? But a teleological theory, just as such, makes no attempt to explain what makes your thought be a representation that grass is green or that aardvarks bark. A prior theory of what (correct) representation is is needed for that. To the shell that is 'teleosemantics', then, it is necessary to add a description of what successful representing, actual representing, is like. The main work of this article is to explain what representing is according to 'biosemantics'.

Above I suggested that the teleologist can take any naturalistic theory of representation at all and turn it into a teleological theory of content. But there is a catch. If the teleologist anchors the notion of function in selection, the theory of representation adopted must allow us to explain how producing inner representations might sometimes benefit an organism. Otherwise it will be a mystery why any organism would contain systems designed by selection to make representations. Surely such a requirement is reasonable, but on careful consideration it turns out to be quite confining.

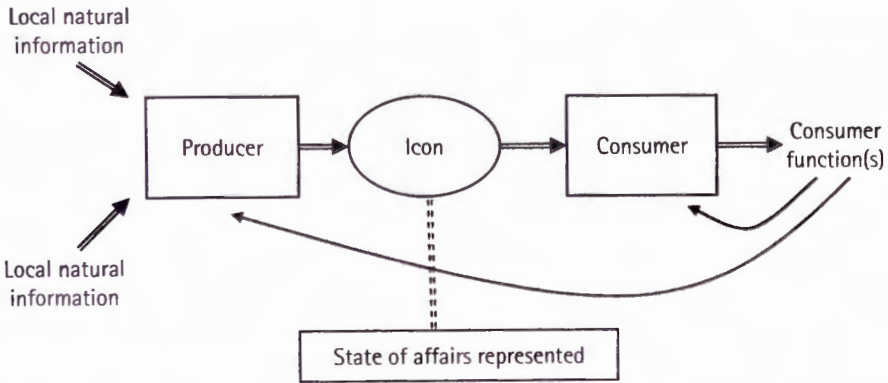
Many naturalistic accounts describe the relation of a representation to what it represents as a simple dyadic relation. This is true, for example, of classical causal or covariational theories, of classical informational theories and of classical picture theories. C. S. S. Peirce, on the other hand, claimed that the representing relation is essentially triadic, involving first the representation (a 'sign'), second something represented, and third an 'interpretant'. If producing inner representations benefits an organism, presumably this will be because the organism uses them in some way. There must be a part of the organism, or some activity of the organism, that understands or interprets these representations. Peirce spoke of the interpretant of a sign as being another sign, but taking this at face value would produce a regress. The interpreter of an inner sign cannot be supposed merely to translate the sign into another inner sign which is again translated, and so forth. 'Interpreting' a sign must ultimately consist in some independently useful activity.

According to biosemantics there are several different kinds of process that use representations. Most theories of representation deal with descriptive representations only—with representations that are designed to represent facts. But directive representations are certainly equally important—representations that tell what to do. And the most primitive and fundamental kind of representation, I believe, faces both ways at once, saying at the same time what the case is and what to do about it. For example, the dance of the honey bee tells where the nectar is and at the same time where the watching bees are to go. I call this last a 'pushmi-pullyu' representation after Hugh Lofting's charming two-headed creature by that name (Millikan 1996). Pushmi-pullyu representations simultaneously describe and direct. Better, since the term 'representation' suggests to many people something more fancy than the simplest examples I have in mind (in particular, it may suggest symbolic forms that are 'calculated over') I prefer to call the inner signs that I describe 'intentional icons' (Millikan 1984)—I will explain why in a minute. Representations that

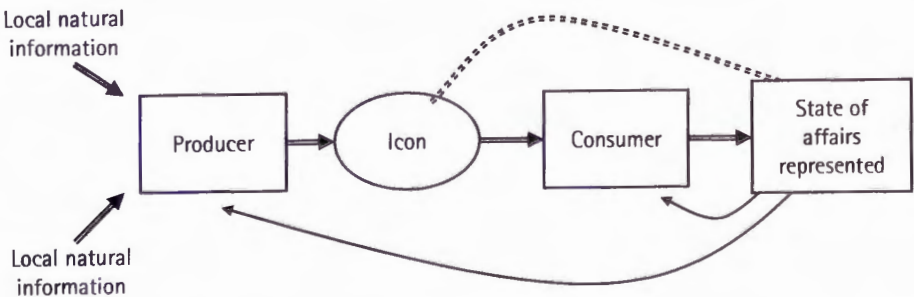
are calculated over—that participate in inference processes—are just one very fancy kind of intentional icon. Below are diagrams of each of the three basic kinds of intentional icon. I will discuss them in turn.

In each of the diagrams there is a producer and an interpreter or ‘consumer’. These have been designed by natural selection or by learning to cooperate with one another. Perhaps each resides in a separate organism; for example, one is part of a dancing bee, the other part of a sister watching bee. Or perhaps they correspond merely to two different functions performed within the same brain. What the producer does helps the consumer to perform functions that loop back to make both the producer and the consumer more likely to survive or to maintain their current settings (selection through learning) or to proliferate. The presence of each is part of the normal mechanism by which the other helps itself to survive or proliferate, and this cooperation is no accident but the result of past selection or learning that operated on both together.

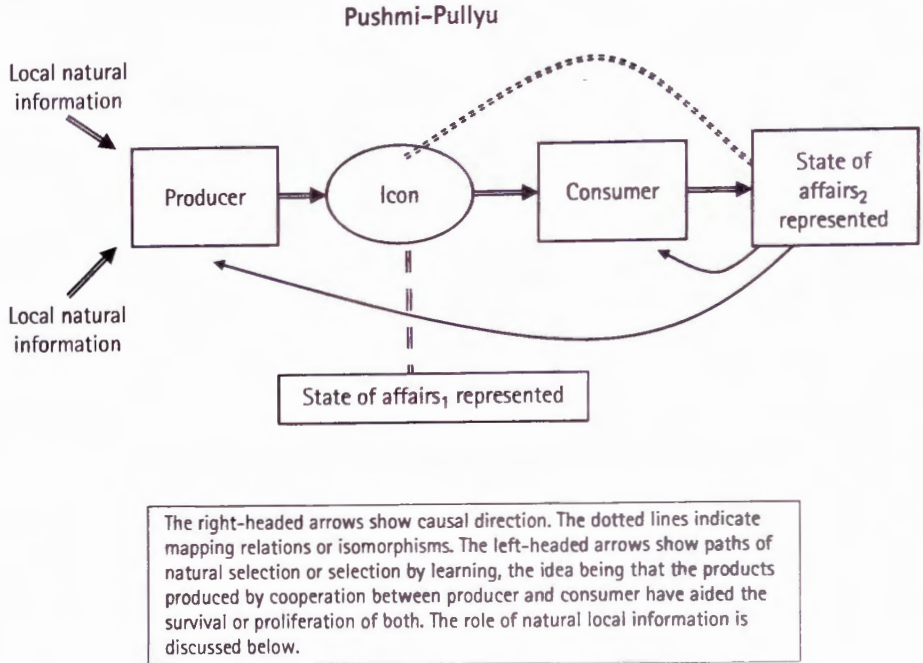
In each diagram the producer produces a sign that will be true or satisfied only if it maps on to some affair or affairs (the plural is for pushmi-pullyu icons) in the world in accordance with certain ‘semantic’ rules. These are rules of correspondence between signs and world affairs that have been instantiated in the past when the consumer and producer or their ancestors have succeeded in performing their



**Fig. 23.1** Descriptive intentional icons



**Fig. 23.2** Directive intentional icons



**Fig. 23.3 Pushmi-Pullyu intentional icons**

cooperative function(s). Consider the bee dance. Suppose that the activities of the dancing and the watching bees fulfil their cooperative function of directing the watching bee to a supply of nectar in the normal way; that is, through the characteristic mechanisms that have accounted in the past for success and subsequent selection of the dance-making and dance-using bee apparatuses. In such cases the dancing bee performs a dance that maps by a certain rule, first, on to the location of nectar. Second, it maps on to the direction and duration of the flight produced in watching bees. In each case, had certain variables in the dance been different—the angle, the speed—the nectar would have needed to be in a different place and the watching bees would have needed to go in a different direction or to a different distance for the dance to serve its purpose in the normal way. The semantic rules of ‘beemese’ are defined by the way in which the set of possible well-formed bee dances maps on to the set of possible nectar locations to determine an isomorphism between these domains, an isomorphism that holds when the bee dance works properly through historically normal mechanisms. Similarly, another isomorphism holds between the set of possible dances and the set of destinations that will result if watching bees find nectar through historically normal mechanisms. The value of a system of representation lies in its productivity, which always depends on an isomorphism between the domain of the signs in the system and the domain of their signifieds. Mappings or isomorphisms are an essential feature of each kind of sign system illustrated. The signs are like abstract pictures. This is why (following Peirce) I

call them 'icons'. *Intentional* icons are produced by systems *designed* to make abstract pictures or icons that will map coincident with predetermined mapping rules to which their consumers are adjusted. When the systems in which they are embedded are functioning normally, they will picture in accordance with these rules, and they are then said to be 'true' or to be 'satisfied'.

I had best add to this that no limit need be placed on the complexity of the semantic-mapping functions that might map intentional icons on to the states of affairs they represent. Isomorphisms can be defined by functions that are as bizarre, as grue-like, as you please. A bizarrely coded secret message from a CIA agent can be as much an 'icon' or 'picture' that maps on to a certain world affair in accordance with a definite semantic-mapping function as any bee dance, sentence, or diagram. Intentional icons must be things apt for use by icon users, but icon users can be very idiosyncratic in their habits. For example, if mental representations are systems of brain happenings or brain states that map on to represented world affairs, no a priori limitation on the kinds of brain happenings or states involved or on the complexity of the mappings employed is implied. Every representation is in some kind of code. The complexity of the code is irrelevant. On the other hand, any intentional icons in the brain would of course have to come with inner interpreters that knew how to read them; that is, interpreters that could be guided by them reliably to fulfil their own functions. Simple codes relying on only a few principles, if they were also highly productive, tapping into rich natural isomorphisms between the domains of the signs and the signifieds, would seem much the most likely to be preferred by natural selection.

The semantic rule associated with a descriptive intentional icon determines a condition or state of affairs that must obtain if the consumer is to perform its tasks, whatever they may be, in the normal way. The consumer varies its activities systematically according to variations in the icons presented to it. The result is that the consumer's activity conforms or is adapted to the condition or state of affairs represented by the icon so that it can perform its functions properly given that condition. Descriptive intentional icons are designed to stand in for world affairs, typically affairs outside the organism or organisms involved, and to vary according to these world affairs, controlling internal or external behaviour as needed to adapt to these affairs. Thus, the bee dance is, in part, a descriptive intentional icon because if the watching bees are to achieve their function of finding nectar by reacting to the dance in the normal way it needs to correspond by a certain rule to a fact about nectar location. (We need not assume it takes any thought on a watching bee's part to react appropriately; so we need not assume that the dance is interpreted by translating it into another sign.) Similarly, consider my belief that there is yogurt in the refrigerator. Which are my belief-using systems? Some, at least, are the systems that make practical inferences and turn the conclusions into practical action. My belief that there is yogurt in the refrigerator can help these systems to perform their function of guiding me into activities helping to fulfil my desires and plans (to eat yogurt, to make breakfast for a guest, etc.) in a normal way (not by serendipitous accident) only if there is yogurt in the refrigerator.

Notice that the content of the descriptive icon is not determined by reference to the kind of tasks that it (with the help of its consumer(s)) normally performs. The icon may be called on to promote a whole series of functions, but each of these will be fulfilled normally only if the icon is 'true'. If the consumer works in part by making inferences, the content is not determined by any particular set of inferences its consumer is disposed to make. It is determined by the fact that *whatever* inferences its consumer makes when functioning properly, the result will be (non-accidental) true belief or successful action only if it *actually* represents according to a certain correspondence rule. This means that its content is not determined by its conceptual role.

Note also that in the case of descriptive icons the producer's job is primarily to make an icon that corresponds by the right rule to a state of affairs. If the producer succeeds in this task, whether through the normal mechanisms of icon production or by freakish accident, the intentional icon is still true. Truth does not rest on whether the means of production was normal. For example, a belief can be true without being knowledge. Taking a classic case first introduced by Dretske (1986), consider the magnetosome of a northern-hemispheric ocean bacterium that normally works by being pulled toward magnetic north, hence toward geomagnetic north, hence away from aerated surface water which is lethal to these bacteria. The magnetosome's pointing in a certain direction is an intentional icon of the direction of lesser oxygen because that is what it needs to correspond to for the bacterium's resultant motion to serve its proper function. If the magnetosome points in the direction of lesser oxygen because, serendipitously, a bar magnet under it is pointed in the right direction, then it tells the truth by accident rather than through normal mechanisms, but it still tells the truth. On the other hand, if a bar magnet attracts the bacterium to its destruction by pointing it towards oxygen, what the magnetosome says is false despite its being produced through a normal mechanism. Attributions of truth and falsity do not rest on whether descriptive icons are produced by normal mechanisms, but only on normal mechanisms associated with their use. (An icon's function cannot be *to have been produced* by something! Functions are effects, not causes.) Of course, Dretske is right that the magnetosome that directs the bacterium in the wrong direction because a bar magnet is held over it is not broken or malfunctioning. In that sense it is functioning 'perfectly properly' (Dretske 1988). But it doesn't follow that it is succeeding in performing all of its functions, any more than a perfectly functional coffee maker is performing all its functions when it is turned on but no coffee has been put in it. Very often things fail to perform their functions, not because they are damaged, but because the conditions they are in are not their normal operating conditions.

Directive intentional icons that correspond to world affairs by their normal correspondence rules are usually said to be 'satisfied' rather than 'true'. The semantic rule associated with the icon determines a condition or state of affairs that its consumers are to produce. The consumer's job is to bring it about that a state of affairs corresponds by rule to the icon. Its job is to 'obey' the producer's 'orders'. (Equally, of course, it is the job of the producer to give orders that will benefit both it and the consumer.) Thus, the bee dance is, in part, a directive intentional icon because it will correspond by a certain rule to the direction of flight of the watching bees

if their dance-interpreting apparatuses succeed in serving their functions normally. Suppose that you have a desire to eat yogurt. The systems designed to be moved by your desires—the ‘consumers’ of your desires—are the systems that make practical inferences eventually turning the conclusions into practical action. If your desire to eat yogurt affects these systems as designed, it will guide them to effect the fulfilment of your desire to eat yogurt. The chief function of a desire is to get itself fulfilled. This is not to say that, on average, desires do get themselves fulfilled. On average they generally get eaten up by bigger opposing desires first, or perhaps no means are known to fulfil them. It is very common for a trait or capacity to have been selected because it sometimes performs a useful function, occasional performance being better than none. The point is that people would not have the capacity mentally to represent various states of affairs as desirable unless these desires were sometimes fulfilled. The capacity would otherwise be useless.

Thus, the cooperation between producer and consumer in the production and use of intentional icons can be accomplished in any of three basic ways. It may be that the producer is the one primarily responsible for making the icon correspond to a certain state of affairs; it may be that the consumer is the one primarily responsible; and in the case of pushmi-pullyu icons it is the responsibility of the producer to make the icon correspond to one kind of affair and the consumer’s job to make it correspond to another. In each of these basic cases, granted the cooperation between producer and consumer comes about through the normal causal mechanisms, the intentional icon will also be a ‘local natural sign’ carrying ‘local natural information’ about the affair or affairs to which it corresponds (Millikan 2004: chs. 3–4). Local natural signs are, in part, abstract pictures of what they represent. Although there is not room to unpack the notion of local natural information here, I mention this because it follows that when intentional systems are functioning normally in accordance with normal explanations, intentional icons represent both by being pictures of what they represent and by carrying natural information as to what they represent. For example, a bee dance is often a local natural sign both of where there is nectar and of where the watching bees will go. In the fundamental sense, *actually* representing involves both picturing and carrying natural information. As such it is not a matter determined by a history of selection. Representing *intentionally*, however, *is* a matter of having a certain kind of history. Also, attributions of truth or falsity and of satisfaction or unsatisfaction make sense only by reference to function, hence by reference to a history of selection. The terms ‘true’ versus ‘false’ and ‘satisfied’ versus ‘unsatisfied’ do not apply to natural signs, the most basic kind of representations.

A common question raised about informational theories of representation concerns how mental representations can carry ‘information about a distant causal antecedent . . . without carrying information about the more proximal members of the causal chain . . . *through which* this information . . . is communicated’, for it seems that such representation skips . . . over (or ‘sees through’) the intermediate links in the causal chain in order to represent . . . its more distant causal antecedents” (Dretske 1981: 158). Similarly, there is a worry about how abstract representations are possible—ones that carry only the information, say, that an object is triangular

and not also that it is isosceles or equilateral. These problems do not arise for the theory of intentional icons. Information carried by normally operating intentional icons is a form of natural information. It does not follow that all of the natural information carried by an intentional icon is carried intentionally. The information that a natural sign carries intentionally is only the information it is its function to carry, the information that its cooperative interpreters know how to use. This information may be very abstract, and it may be about very distal affairs. If its consumers are so designed that they can use only the information that something is triangular, then that is all the information that the icon carries intentionally. If they are so designed as to use only the information that a predator is near, then the intentional icon will not intentionally carry information about any more proximal affairs, such as patterns on the retina.

Similarly, not every stimulus that an organism discriminates on the way to producing an intentional icon is represented intentionally. Nor must an organism be capable of infallibly discriminating the distal objects, properties, or kinds that it intentionally represents from all others that are similar. It needs only a fallible capacity to recognize some natural signs or other of these things under some local conditions. Possibly it even gets things wrong a large part of the time because a large part of the time supporting conditions on which its mechanisms of icon production rest are absent. Similarly, the rabbit's danger thump may be elicited more commonly when rabbit danger is absent than present. (What matters is the converse—that when danger is present it should usually be elicited.)

The basic theory of representing on which biosemantics rests is a picture theory and an informational theory, but equally it is a functionalist theory. The basic idea is that what makes something into a representation, for example into a mental representation, is not, of course, what it is made of, but what functions it performs and/or *how* it performs these functions. Items that function in certain ways are representing, and if they have been designed to function in these ways they are representing 'intentionally'. They are representing, that is, in accordance with natural purposes and such that they can be said to be true or false, satisfied or unsatisfied. (The intimate connection of functions that have been selected for with purposes is argued in Millikan 1984; 2004: ch. 1.)

According to biosemantics, basic mental representations always represent complete states of affairs. Mental terms are not endowed with meaning first and then used to build mental sentences. 'What makes the mental term "horse" stand for horses?' is not the place to begin. Parts and aspects of complete intentional icons represent parts and aspects of complete states of affairs only as abstracted from completed icons. This general point is most apparent, perhaps, with the simplest and most common cases of pushmi-pullyu intentional icons, exemplified by nearly all animals' signals and also by the ubiquitous chemical signals running in the bloodstream that direct responses from various organs and cells. These signals, taken along with their times of occurrence and sometimes with their places of occurrence, are intentional icons because variations in the times and places of occurrence correspond to variations in the times and places of the complete affairs represented. For example, the time and place of the

mother hen's food call to her chicks descriptively shows the time and place of food and directly tells the time and place her chicks are to come. But it is evident that there is no sense in which a particular time stands for itself or a particular place for itself outside the context of some such signal. Similarly, in telling what direction the bacterium is to go the direction in which its magnetosome points stands for itself, but there is no prior sense in which a direction stands for itself.

The biosemantic account also implies that there are not and could not be intentional icons that lacked attitude. Rejected is the Fregean idea that first a proposition is represented, then an intentional attitude added. Intentional icons always have, as such, functions and their functions automatically create attitudes. Hypothetical thinking, for example, or just thinking of possibilities, is an extremely sophisticated activity, and one that is only possible for a creature that sometimes uses the results in the production of ordinary descriptive and directive representations. It is because thoughts 'of possibilities' have the function of sometimes turning into more basic kinds of representations that they exist as representations at all. Similarly, desires are intentional icons only because they are designed to turn, under certain conditions, into full-blown intentions, whose functions are to effect their own fulfilment more directly. There would be no benefit in the capacity to have desires if desires did not sometimes travel the whole route through intention into action.

Intentional icons represent complete states of affairs. This implies that they represent not only properties but also the things that have those properties. When produced normally, intentional icons also carry natural information corresponding to what they represent. This implies that there can be natural information as to what things have what properties—including what *individuals* have what properties. Contrast Dretske's description of the natural information carried by signals. As Dretske describes the matter, although a signal can carry the information that an individual *x* is *F*, there is no part or aspect of the signal that carries the information that it is *x* that is *F*. For example, the petrol gauge on your car may carry the information that your petrol tank is half full, but no aspect of the signal indicates *which* tank is half full (Dretske 1981, 1988). You have to know that independently. The petrol-gauge reading does not represent its subject, nor could it on Dretske's theory of natural information, because there are no natural laws that pertain to any individuals just as such. A necessary and central feature of the theory of *local* natural information (Millikan 2004) is that it explains how a natural sign can signify which individual it carries information about. The result is an explanation of how intentional representations of individuals are possible, something for which, to my knowledge, no other naturalized theory of intentionality accounts.

A common question raised about the programme of biosemantics is how representations such as human beliefs and desires, which in numerous instances are entirely unique to the individual who has them, can have acquired functions through a process of selection. In outline the answer is straightforward. Compare the design of a camera. A camera is not designed to take any particular picture, but to vary the picture it produces depending on the scene in front. If a particular pattern is in front, it will function properly if it produces a likeness of that pattern. Similarly, of course,

one's eyes are not designed to see any particular object but, roughly speaking, to see whatever object lies in front of them. If a particular person is in front of your open eyes, it is a function of your eyes to help produce an accurate perception of that person. Similarly, the function of an adding machine is not to give any particular answer, but to give the sum of the numbers put into it. Natural selection has designed cognitive systems not to turn out particular products, say particular beliefs and desires, but to turn out quite different beliefs and desires depending on environmental circumstance. In the case of human beliefs and desires, however, the matter is considerably more complicated than with the camera. In order to turn out beliefs that will vary depending on states of affairs in the environment and in order to tune the systems that use these beliefs during practical and theoretical deliberation and in the production of useful action, humans must first develop adequate empirical concepts. Indeed, to complete the biosemantic programme a rather long story needs to be told about conceptual development. We must explain how the producers and consumers of beliefs and desires can learn or be tuned to employ empirical concepts cooperatively without actually practising together through the production of concrete actions. We must explain how their representation, production, and use dispositions can be tailored in advance to fit one another. That story is told in Millikan (1984: chs. 15–19; 2000: esp. ch. 7; 2004: p. IV, esp. ch. 19).

Many critical questions about the biosemantic theory first presented in 1984 and 1989 have come and gone, but there are three that have been especially tenacious. I will say a word about each.

What an intentional icon represents descriptively is an affair to which it must correspond if its consumers are to perform their functions by normal mechanisms. They will perform their functions through normal mechanisms only if external conditions are such as to allow these mechanisms to operate properly. Taking for her example the female-hoverfly detector in a male hoverfly's visual system, Karen Neander (1995) has objected that among the external conditions needed for the detector's consumers to perform all their functions are that the female is fertile and that she won't be eaten before she reproduces, hence that on the biosemantic theory these facts about the female must be part of what is represented by the detector in the male's visual system. What this overlooks, however, is that an intentional icon must also have a producer and that it must be a function of the producer to make an icon that corresponds to the condition it represents. If the producer has a function there must be a normal mechanism by which it performs that function. This, however, would require the male hoverfly's visual systems to be sensitive to natural signs of fertility in female hoverflies and of liability not to be eaten. But on no theory of information, certainly not on the theory of local natural information, does the male hoverfly use or even encounter any such natural information (Millikan 2004: ch. 6).

A second question concerns the possibility of biological systems whose jobs are to produce false representations. For example, people who are overconfident may be more successful at performing certain tasks than people who evaluate their skills correctly. Notice first, however, that it will not be the falseness per se, but only the high confidence that contributes to success. If one is completely and perfectly competent

at a task, there certainly will be no gain in believing one is not! Notice second that there are many biological systems that ride piggyback on systems developed earlier for quite different purposes. If there actually were systems whose jobs were to distort certain beliefs they would have to ride on more general systems whose basic jobs were to produce true beliefs. Otherwise there would be no semantic rules in place according to which the distorted beliefs would be false. The various concepts out of which any beliefs are formed are designed to serve purposes in arbitrary belief contexts. The systems responsible for concept development tune these concepts and the systems that normally use them for general purpose use, not for any one specific use such as increasing one's confidence. The semantics of mental representation is productive. That is what the 'picturing' or 'mapping' guarantees.

Third, consider Pietroski's tale about the kimus and the snorfs (1992). The kimus are attracted by the red sunrise glowing over their local mountain so that they climb up it each morning. Thus, they conveniently avoid their chief predators, the snorfs, who pass by each morning below. Moreover, this is how the attraction to red light got selected for in kimus. Those not attracted by red light got eaten. On the biosemantic view, Pietroski claims, 'kimus climb the hill *because they believe* the hill is snorf-less', and when they approach red things that are not snorfs 'they are *acting on the belief* that the area in question is snorf-free' (1992: 276). Given that kimus 'can't reliably discriminate snorfs from non-snorfs', (*ibid.*), it is implausible, he claims, that the kimus have any beliefs about snorfs. In summary, on the biosemantic account

[a] system can have the belief that P is instantiated without having *any* systematic ability to tell whether P is instantiated (in a given region at a given time). Indeed, instantiations of P can be completely irrelevant to the system's tokening of the belief that P is instantiated. The corresponding intentional explanations of such a system's behaviour will . . . be very implausible.

(Pietroski 1992: 268)

What Pietroski describes in the kimus seems to be a simple tropism. They are attracted to red the way a moth is attracted to light. Apparently they have neural pushmipullyu intentional icons or signals that tell where the snorfs are fewer and hence where to go. *Of course* these icons are not at all like beliefs. Beliefs are formed only after the acquisition of concepts, which generally rest on multiple ways of recognizing. Further, the functions of beliefs involve participation in inference. (For a discussion of empirical concepts and of inference see Millikan 2000.) Pietroski seems to assume that an 'intentional explanation' of an animal's behaviour must not only be a belief-desire explanation but must also be a straightforward causal explanation. Why intentional explanations are causal but not straightforward is explained in Millikan (1993, 2007). Moreover, full intentional explanations do not begin with the presence of intentional representations, but explain also how the intentional representations get formed. The red light is definitely involved in an 'intentional explanation' of how the kimus avoid snorfs.

Finally, there are no *distal* objects or stimuli that *any* organism has the capacity to discriminate under *all* conditions. All successful discrimination of distal affairs

depends on merely local natural information. Local natural information rests on correlations that are not perfect but that are not accidental either, for they must persist throughout a spatio-temporal region for a reason (Millikan 2004: chs. 3–4). On this analysis the kimus do get local natural information each morning about the direction of fewer snorfs. Similarly, although there is no causal connection, the correlation between magnetic north and lesser oxygen used by the anaerobic bacteria persists in the northern hemisphere for a reason. It carries local natural information about the location of lesser oxygen. The intentionality that characterizes pushmi-pullyu icons responsible for simple tropisms of this kind is the limiting case of intentionality. It is intentionality in the way zero is a number. If your theory doesn't count in these cases you will find that it fails to account for any of the obvious cases either.

## REFERENCES

- Dretske, F. (1981), *Knowledge and the Flow of Information* (Oxford: Blackwell).
- (1986), 'Misrepresentation', in R. Bogdan (ed.), *Belief: Form, Content, and Function* (Oxford: Clarendon), 17–36.
- (1988), *Explaining Behaviour* (Cambridge, Mass.: MIT Press).
- Millikan, R. G. (1984), *Language, Thought, and Other Biological Categories* (Cambridge Mass. MIT Press).
- (1989), 'Biosemantics', *Journal of Philosophy*, 86: 281–97.
- (1993), 'Explanation in Biopsychology', in J. Heil and A. Mele (eds.), *Mental Causation* (Oxford: Oxford University Press), 211–32; repr. in Millikan, *White Queen Psychology and Other Essays for Alice* (Cambridge Mass.: MIT Press); and in C. Macdonald and G. Macdonald (eds.), *Philosophy of Psychology: Debates on Psychological Explanation* (Oxford: Oxford University Press, 1995), (pts VII–X).
- (1996), 'Pushmi-pullyu Representations', *Philosophical Perspectives*, 9: 185–200; repr. in L. May and M. Friedman (eds.), *Mind and Morals* (Cambridge, Mass.: MIT Press, 1996), 145–61.
- (2000), *On Clear and Confused Ideas* (Cambridge: Cambridge University Press).
- (2002), Biofunctions: Two Paradigms, in R. Cummins, A. Ariew, and M. Perlman (eds.), *Functions: New Readings in the Philosophy of Psychology and Biology* (Oxford: Oxford University Press), 113–43.
- (2004), *Varieties of Reference: The Jean Nicod Lectures 2002* (Cambridge Mass.: MIT Press).
- (2007), 'An Input Condition for Teleosemantics?', *Philosophy and Phenomenological Research*, 75, 436–455.
- Neander, K. (1995), 'Misrepresenting and Malfunctioning', *Philosophical Studies*, 79: 109–41.
- Pietroski, P. M. (1992), 'Intentionality and Teleological Error'. *Pacific Philosophical Quarterly*, 73: 267–82.