PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Explanatory Role of Information [and Discussion]

Fred Dretske, A. Clark, Y. Wilks, D. Dennett, R. Chrisley and L. J. Cohen

Phil. Trans. R. Soc. Lond. A 1994 **349**, 59-70
doi: 10.1098/rsta.1994.0113

The explanatory role of information

By Fred Dretske

Philosophy Department, Stanford University, Stanford, California 94305, U.S.A.

The mind can be viewed as an information-driven control system. To make this work, the idea of information must be operationalized in such a way as to give semantic properties (meaning, content) a role in the explanation of system behaviour. This can be achieved by exploiting a statistical concept – mutual information – from communication theory. On this interpretation, some of the behaviour of information-driven control systems is causally explained by the statistical correlations that exist between internal states and the external conditions about which they carry information.

1. Introduction

It is useful to think of the mind as a control system. Aaron Sloman thinks of it that way, and so do I. It is hard to see how else to think about it. What good is a mind if it does not do something?

That is the easy part. The hard part is to say what kind of control system the mind is. What is special about it? Everything that does anything – animate or inanimate, natural or artificial – has something inside it that causes and, in this sense, controls its behaviour. (In Dretske (1988) I distinguish behaviour (doing something) – for instance, moving your arm – from the events (e.g. arm movements) that result from such behaviour. Arm movements can be caused by wholly external events, but if I move my arm (which I term behaviour), there must be something inside me that causes my arm to move.) Maple trees shed their leaves in the autumn. That is the sort of thing maple trees do. This behaviour is controlled by chemical processes occurring inside the tree. Just like our own cognitive states, these processes are sensitive to external affairs – temperature, light, and so on. The same is true of artefacts. The behaviour of an alarm clock – whether it rings and when – is determined by the internal arrangement of its parts. However, neither oak trees nor alarm clocks are minded. They do things, of course, and the things they do are controlled by internal processes (some of which carry information about external affairs), but nothing they do is explained by what they believe and desire. They do not have beliefs and desires. They are not minded. Their control systems are not of the right sort.

Our control system is of the right sort; or so we all think. Not only do we have beliefs and desires, hopes and fears, purposes and plans; some of what we do is explained by these residents of the mind. We call such explanations – those couched in terms of beliefs, desires, purposes and intentions – our reasons for doing what we do. That is what makes the idea of the mind as a control system

Phil. Trans. R. Soc. Lond. A (1994) **349**, 59–70
Printed in Great Britain

an attractive idea. The problem with this way of thinking about the mind, though, is distinguishing beliefs and desires, purposes and intentions, the alleged mental bases of action, from the sort of events controlling the behaviour of clocks and trees. Indeed, what is it that distinguishes the reasons we have for doing what we do – moving our hands, say – from the electrical, chemical and mechanical events that make the hands move?

One way to approach this question, a way that I have pursued for the last fifteen years, is to think of the mind as an information-driven control system. If the mind is ever really in control (and some will surely doubt this), then what distinguishes this control system from others is that it deals in a different causal currency. Something like meaning or content – what it is we believe and desire, what it is we perceive and intend – pulls the levers and turns the gears. What makes me move my hands (when I do so intentionally) are facts about what I perceive and what I want; facts relating to the content or meaning of my internal states. This is something altogether different from the pushes and pulls that explain why clocks move their hands. On this way of thinking about the mind, minded systems are semantic engines: meaning is the fuel. Unlike oak trees and clocks, what explains the behaviour of minded systems, at least when they act for reasons, is not simply internal events that have meaning, but the fact that they have meaning.

If one is a realist (which is not the same as a dualist) about the mind, one is virtually driven to this conception of the kind of control system the mind is. If the mind controls behaviour, if we ever do things for reasons, then it is surely what we think and intend, what we desire and prefer, the meaning or content of our mental states, that is the explanatorily relevant fact about them. Why bother thinking that your keys are on the desk, not in your trouser pocket, if what you think – that the keys are on the desk – is not somehow implicated in your looking on the desk?

It is this conclusion, an inescapable one for realists about the mind, that has made a good many contemporary philosophers skeptical about the mind. At least they are skeptical about the possibility of viewing the mind as a control system. The consensus seems to be that there can be no semantic engines, nothing whose behaviour can be causally explained by the meaning (if any) of its internal states (see, for example, Dennett 1983). Meaning, including what it is people think and want, just isn't the sort of thing that could control the trajectory of a 170 pound object as it looks for its keys. We know enough about how the body works to know that the muscles are controlled by electrical signals emanating from the central nervous system. These impulses, or the events that trigger them, may have a meaning, they may carry information, but it is not this meaning or information that explains their effect on bodily movement. We are syntactic engines. The fuel is, as it always is, the stuff of physics and chemistry. If the way the mind is supposed to exercise control is via information, then, quite simply, the mind cannot be in control. From an explanatory standpoint, the mind is irrelevant – epiphenomenal, as philosophers like to say.

I see this as something of a dilemma for people who think of the mind as an information-driven control system. If the mind is to be made relevant causally to what we do, and the mind is, at the same time, understood in terms of abstractions like information, meaning and content (if those are the currency with which mental affairs are transacted), then it is difficult to see how the mind can make a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

causal contribution. What does the mind do, and, even more baffling, how does it do it? If information is the fuel, then what is the spark that will release its power?

To understand the dynamics of a semantic engine, one whose output (or some of whose output) is powered by information, one must first understand something about the nature of the fuel.

2. Objective information

There is a statistical conception of information, derived from communication theory, that defines amounts of information in terms of the conditional probabilities between events occurring at a source (S) and receiver (R). Call this quantity mutual information. The mutual information between R and S is a measure of the statistical dependence of events occurring at these two places. The greater the dependence between events occurring at R and S, the more mutual information there is between these points.

If signals are understood to be information-carrying events (sounds, gestures, electric currents), then communication theory is concerned with the statistical properties of these events, not with the information they happen to carry in any ordinary sense of the word 'information'. My dictionary tells me that what we ordinarily mean by 'information' is something like knowledge, data, news or intelligence. Information, in the ordinary sense, and in the sense it bears in much of cognitive science, has to do with the message. Communication theory, on the other hand, is concerned with the medium. It is concerned with how probable or improbable the events that carry information are, not with what, if anything, these events say or mean. For this reason it is often said that communication theory is not really a theory about information at all; certainly not about the sort of merchandise, call it semantic information, that we are commonly concerned with when we use this word. There is some truth in this charge: mutual information is a statistical quantity and it must be carefully distinguished from the news, intelligence or message, that an event, however probable or improbable it might be, happens to carry.

Nonetheless, although the chief concern of communication theory is the statistical properties of signals and channels, not the semantic information (if any) that these signals happen to carry over the channels, the statistical properties turn out to be relevant to what semantic information a signal can carry. Unless there is a statistically reliable channel of communication between S and R, the signals reaching R from S cannot indicate what is happening at S. A broken voltmeter, whose pointer (R) does not bear the right statistical relations to the voltage (S) that it is supposed to measure, cannot convey semantic information about the value of that quantity. It cannot say what it is supposed to say, what it was designed to say; namely, that the voltage is so-and-so. Using such an instrument, one cannot learn or come to know what the voltage is. This is why one cannot tell the time of day from a broken clock, even if it happens to register the correct time. This is why one receives little or no information from unpredictable liars, even when they happen to be telling the truth. Hence, communication theory, though not directly addressing semantic topics, does, it seems, formulate a plausible objective condition on the flow of semantic information. No signal can carry semantic information (the information that P), unless the channel over which

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

the signal arrives satisfies the appropriate statistical constraints of communication theory.

Communication theory can be interpreted as telling one something important about the conditions that are needed for the transmission of information as ordinarily understood, about what it takes for the transmission of semantic information. This has tempted people (such as myself) to exploit it in semantic and cognitive studies, and thus, eventually in the philosophy of mind. Perhaps then, the semantic information in a signal about a source and, therefore, what a signal says about a source are simply the conditions about which signals bear appropriate amounts of mutual information.

This is not the place for contentious details (and most of the details certainly are). Suffice it to say that, from the point of view of cognitive science and the philosophy of mind, the basic extractable idea in communication theory is that semantic information is constituted by the network of dependency relations existing between two places. Just how reliable this connection must be for the receiver to obtain semantic information about that source, information that a particular condition exists there, I will not try to say here (see Dretske 1981). Perfect reliability, modulo some set of existing circumstances, certainly seems enough. If, under conditions C, events at one point depend on events at another in such a way that R would not occur at the one unless an event of type S exists at the other, then in these circumstances, tokens of R will carry information that an event of type S exists at the source. Perhaps something less than perfect reliability (modulo conditions C) is enough, but the basic idea of communication theory, that of conditional dependence or correlation, is always at work when we speak of information. This is why we think of gauges as carrying information about the quantities they are used to measure. The gauge tells us that we are low on gas because there is, or so we believe, a reliable enough correlation between what it registers and the amount of gasoline left in our tank. This is why we think of fingerprints as carrying information about who handled the gun; why books, newspapers and television convey information about matters not directly observed. There is, to some degree, a dependency of the conditions which we can observe on the page or screen, on what is taking, or has taken, place elsewhere. This, too, is why it is so natural to think of the brain as obtaining, by means of the senses, enormous quantities of information about current events, which (with the aid of memory) it can later retrieve and use. In fact, this is precisely what the senses do, what they are for: they make what happens inside the skull depend on what happens outside the skull, thus providing the brain with the information it needs to coordinate behaviour with the conditions on which the success of behaviour depends.

I will not try to defend this conception of information here. I merely note that it makes information into a perfectly objective commodity, something that is independent of our interpretive stances and epistemological attitudes. On this view, whether there is information at point R about point S does not depend on whether anyone knows or understands this fact. It does not depend on whether anyone obtains or uses the information. The relation between temperature and volume that makes the mercury in a thermometer say or mean that the temperature is 85 is a perfectly objective fact about metals that does not depend, for its existence, on our knowledge or understanding. Expanding metals meant what they did, and thus carried information about temperature, long before anyone re-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

alized this and exploited it to build thermometers. For the same reason, whether or not a brain receives sensory information about the outside world does not depend on whether that brain, or any other, knows it is getting information. You can, in this sense of information, get information without getting the information that you are getting it.

This sense of the word information is very close to what Paul Grice (1957) called natural meaning. This is the sense of 'meaning' in which shadows in the east mean the sun is in the west, dark spots on the X-ray mean there are cavities in the tooth, and wet streets mean that it has been raining. It is the way natural signs (footprints, clouds, rock formations, symptoms of a disease, and tree rings) mean or signify something about their causal origins. This is the sense in which smoke means fire – not the word 'smoke' (for it means smoke, not fire) but smoke itself. (The sense in which the word 'smoke' means smoke is quite different from the way the sound of a smoke alarm means smoke. The first is what Grice calls non-natural meaning, a sense of the word in which something (e.g. 'there is smoke') can mean that there is smoke even when there is no smoke. In the natural sense of meaning, nothing can mean that there is smoke unless there is, in fact, smoke.)

We also speak of one condition indicating something about another. I go to the door when I hear the doorbell because I take its ring to indicate, mean, carry the information that, someone is at the door. I shall use these terms – information, indication and natural meaning – interchangeably.

Some people, I know, like to think about information as something like knowledge or significance. Nothing can carry the information that P unless it signifies P to or for someone. If no one regards the dark clouds as indicating (carrying information that) a storm is approaching, then the clouds do not indicate, do not mean, do not carry the information, that a storm is coming. All indication, all information, all natural meaning, requires an understanding of what is indicated, of what information is being conveyed, of what is meant. In a world devoid of intelligence and understanding, there is no information.

I cannot legislate for use. If someone wants to speak about information in such a way that nothing can be information unless it is information for or to someone in this way, then they are free to do so. I think it is confused (in the same way the identification of true with true-for-me is confused), but I do not have the time to argue about it here. I merely note that such a notion of information is useless as an analytic tool in semantic and cognitive studies. It is useless because it presupposes the very ideas – meaning, thought and knowledge – that we need a concept of information to clarify and explain. There is no progress, either scientific or philosophical, in developing an information-based theory of the mind if the information we use to explain the mind is already tainted with what we are developing a theory to understand.

The second point I wish to emphasize about this objective notion of information is that information, so understood, is an extrinsic, relational property of events. Information is certainly a perfectly objective thing, but objective in the way that being a photograph or a dollar bill is objective. Two pieces of paper can be indistinguishable and yet, one may be a real dollar bill and the other counterfeit, or one a photograph of Jim and the other a picture of Martin. This is possible because, as we all know, what makes a piece of paper real money or a photo of Jim has to do with its relations, primarily causal, to other things. What is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

in my wallet is (I think) real money because it was printed in the right place by the right people. What makes a piece of paper a picture of Jim (and not his twin brother Martin) has to do with the fact that Jim (not Martin) was the causal source of the reflected light that exposed the film from which the print was made. For the same reason, whether event A carries information and, if so, what information, depends not on the intrinsic properties of A, but on its relations to other events. Disconnect your speedometer. It will no longer supply information about how fast your car is going. Yet it will register exactly the same as what a properly functioning speedometer registers in a stationary car. Just as you cannot tell, simply by looking at the paper, whether it is a real dollar bill or a perfect counterfeit, whether it is a picture of Jim or a picture of Martin, so you cannot tell, by looking at the speedometer, whether it carries information or, if it does, what information it carries. This, incidentally, is why you cannot look inside the skull and expect to see what information is being processed there or even whether information is being processed there. The events that carry information are inside, but the information they carry is not.

3. Can information have causal effect?

What I have just sketched is a way of operationalizing the concept of semantic information. What it gives us is an objective commodity, relational in nature, that comes reasonably close to what scientists and laypeople are talking about when they talk about information. It is the sort of thing – reliable dependency – that instruments, gauges and meters deliver about the quantities they are designed to measure. It is the sort of thing that we all expect to find, to some degree, in the books, newspapers and magazines on which we rely. It is the sort of thing that sense perception makes available to the control mechanisms of living systems, the sort of thing that, once it is obtained, can be stored and retrieved for later use. If the mind is a control system, this, surely, is the stuff on which, and by means of which, it exercises its authority.

There is, unfortunately, a fly in this ointment, which has been buzzing, annoyingly, and with increasing volume, for the past twenty years. The people it has been annoying are people like me who seek, in the idea of information, some basis for a scientific study of the mind. For it is one thing to say, as I have, that, in this objective sense, information is picked up in perception and stored in memory. This says little more than what we all already believe: that the senses, when operating normally, institute systematic dependencies (that may be preserved over time) between what happens inside the head and what happens outside the head. This much, I hope, is uncontroversial. It is, however, quite another matter to say, as I would like, and as one must say if one thinks of the mind as an information-driven control system, that such information is used in the control and guidance of behaviour. For this attributes an important new quality to information, one not so far explicitly acknowledged. It attributes causal effect, explanatory relevance, to information, and it is not at all clear that information, as we have operationalized it, has this essential power.

Let me illustrate this problem by reminding you of the broken speedometer I mentioned earlier. From the point of view of the instrument that carries information, and thus from the point of view of all causally downstream processes, broken speedometers are indistinguishable from functioning speedometers. An

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

instrument that says the car is going 0 mph, thus delivering (say) six bits of information, is indistinguishable from a broken instrument that delivers no bits of information in registering the same. This result is a consequence of regarding information in a relational way, as constituted, not by any intrinsic fact about the signal (pointer registering '0') but by the set of relations between that signal and the conditions (in this case, car speed) about which it carries information. Because there is absolutely nothing about an information-bearing signal to distinguish it from an information-barren signal, there is a problem of how to make the information in a signal do any causal work. How can the fact that the speedometer carries the information that the vehicle is stationary explain anything when, if it (being broken) carried no information about the speed of the vehicle, it would (or might) register exactly the same?

Put that speedometer, or its informational counterpart, in the head of an animal, and you have the problem of how to make information do any real work in the control of animal behaviour. If we think of certain control systems (unlike those in clocks and oak trees) as consumers and users of information – and of their owners, therefore, as genuine semantic engines – then we must, it seems, think of these control systems as sensitive to information. We must think of them as sensitive, not merely to the physical properties of the signals that bear information (a needle pointing at '0'), but to the information it bears (the fact that the car is not moving). However, if information does not exist *in* the head, but in that set of relations between head and world, how is it possible to make control processes in the head respond to it? It seems an impossible task.

So it seems to many philosophers. Causality, these philosophers (e.g. McGinn 1989) like to say, is a local affair. Objects (and this includes information-bearing signals) that are physically indistinguishable, that (like our two speedometers) share all their intrinsic properties, cannot differ causally. Information, on the other hand, is a relational affair. It is not local. As our two speedometers show, physically indistinguishable situations can differ informationally. This being so, the conclusion seems inescapable: information is causally irrelevant. It is not the fuel that powers the engine. There are no semantic engines and there cannot be. There are only syntactic engines, systems that are powered by the physical events that carry information. These events doubtless carry information, but, from an explanatory standpoint, for purposes of understanding why the system behaves the way it does, we need never mention this fact. All that is relevant to the way the system behaves is whether the internal needle points at '0', not whether this event carries information about the speed of the vehicle.

Philosophers have devised various strategies for living with this unpalatable conclusion: instrumentalism, computationalism, eliminativism, and so on. These are all ways of conceding that, contrary to our ordinary ways of speaking, meaning plays no causal role in the behaviour of living systems. Literally speaking, no one ever goes to the fridge because they want a beer and think there is one left in the fridge. We certainly talk this way, and we can, perhaps for practical (including predictive) purposes, go on talking this way, but, strictly speaking, it just is not so. Despite a widespread conviction that it happens all the time – if not with beers in refrigerators, then with other objects in other places – it cannot happen. It cannot happen because, such descriptions, if taken literally, imply that meaning or content, what it is we think and what it is we want, causally explains what we do. Meaning just does not have that kind of power.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Personally, I prefer to reject, rather than live with, this conclusion. I think there are, quite literally, semantic engines. That is what we are. I think we are semantic engines because I think we have, inside our skulls, information-driven control systems. That is what we call our mind. I reject the idea that there can be no such thing, not because I reject the premises used to reach this conclusion, but because I do not believe the premises imply this conclusion. I agree that, in the relevant respects, causality is a local affair. I also agree that information and meaning are relational. But none of this implies that information is causally irrelevant to what we do. None of it implies that behaviour cannot be explained, causally, by the kind of relational facts that constitute meaning. To suppose it does, I submit, betrays a muddled conception of the different ways behaviour is causally explained.

4. Types of causal explanation

Causal explanations, as we all know, are context sensitive. What we pick out as the cause of an event depends on our interests, our purposes and our prior knowledge. Almost any event, E, depends on a great variety of other events in such a way as to make any one of them eligible, given the right context, for selection as the cause in a causal explanation of E.

In addition to context sensitivity, there is another source of variability in causal explanations. In seeking an explanation for an event – whether the event in question is a piece of animal behaviour or something else – there is what I have elsewhere (Dretske 1988) called the triggering causes of an event and what I have called its structuring causes. For present purposes I do not think I need technical definitions. A few examples should suffice to make the distinction clear enough to apply.

An operator moves the cursor on a screen by pressing a key on the keyboard. Pressure on the key makes the cursor move. Pressure on the key is what I call a triggering cause of movement. If asked to explain why the cursor moved to the left, it would be natural to mention the fact that the operator pressed the backspace key. Sometimes, though, in explaining why the cursor moved, we are more interested in standing conditions, actual electrical connections in the computer or possibly the software that assigns functions to the various keys. This is especially so when cursor movement (in response to pressure on the key) is unexpected or unusual. Imagine a puzzled operator, watching the cursor move while poking the key, asking: 'Why is the cursor moving?' Because the operator knows that pressure on the key is making the cursor move (that, in fact, is what seems puzzling), a different explanation of cursor movement is being sought. The operator is looking for what I call a structuring cause: in this case what brought about or caused the machine to occupy a state or to be in a condition in which pressure on that key has this effect. It is known, or is easily assumed to be known (after a few presses on the key) that M (cursor movement) is being caused by C (pressure on the key); what the operator wants to know is why it is. Who or what made M depend on C in this way?

A terrorist plants a bomb in a general's car. The bomb sits there for days until the general gets in the car and turns the key in the ignition. The bomb is detonated and the general is killed. Who killed the general? The terrorist, of course. How? By planting a bomb in the car. Although the general's own action,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

turning the key in the ignition, was the triggering cause of the explosion, the terrorist's action, something that took place days before, was the structuring cause of the explosion. Structuring causes, cause triggering causes to have the effects they do and, thus, serve indirectly as causes of those effects.

Triggering and structuring causes, though always distinct, may sometimes appear to 'fuse'. Imagine a dim-witted terrorist forgetting where the bomb had been planted. A few days later, when in need of a car, the terrorist is blown up while stealing the 'wired' car. Is the terrorist both the triggering and structuring cause of his/her own death? He/she certainly created the conditions that enabled the (unintentional) explosion, but that only means that one and the same person was involved in both causes. It does not show that the causes are the same. What was done to trigger this result is different from what was done to structure it. It was turning on the ignition that triggered the explosion and wiring the bomb to the ignition (a week ago) that structured it. Though both events (in this case actions) involve the terrorist, they are quite different.

Just as different actions of the same person can operate as triggering and structuring causes of the same event, different states of the same object can function in the same way. An object possessing one property might be the triggering cause of an event, whereas its possession of a quite different property can be the structuring cause. It is important to keep this possibility in mind when thinking about events occurring in the nervous system and the way these events control behaviour. For there is reason to think that, with respect to some behaviour (that which we regard as deliberate or intentional), it is the brain's possession of certain intrinsic properties, electrical and chemical ones, that triggers bodily movement, whereas it is the brain's possession of certain extrinsic, informational properties that structures bodily movement.

With this distinction between triggering and structuring causes in mind, let us look at a few examples to see how it applies to explanations of intentional behaviour. I begin with the behaviour of artefacts and shall consider animals, those with minds, shortly.

5. Explaining behaviour

You bought an appliance – I shall call it Gizmo – in kit form and you carefully assembled it by following an elaborate instruction manual (example taken from Dretske 1992). When Gizmo is finished, you plug it in, push the start button and wait for it to perform. Instead of whirring into action, as you had been led to expect, there is a humming noise and smoke begins to rise from the transformer. You quickly turn it off, wait a few minutes, and try again. The same thing happens. Understandably, you want to know why Gizmo is behaving in this strange way.

Well, in one rather standard, but, in this case, totally inappropriate, way of thinking about causal explanations, the causal explanation for Gizmo's disappointing behaviour is obvious. Gizmo hums and emits smoke whenever you push the start button. Pushing the start button is, therefore, probably the cause of its behaving that way. Stop pushing the start button and it will not behave that way anymore. Try it and see.

This is an inappropriate way of thinking about Gizmo's behaviour because, given the context I constructed, it was already obvious that pushing the start

button, turning the device on, was causing it to behave that way. So that answer is not the answer to the question. That explanation cannot be the explanation we are looking for because that fact was already obvious when we asked for an explanation. So even if turning Gizmo on is a causally relevant condition for its behaving that way, this is not the causally relevant condition we are looking for in our search for a causal explanation of Gizmo's behaviour.

In seeking an explanation for Gizmo's behaviour, what we are looking for, of course, is an explanation of what went wrong (assuming, for the moment, that something did go wrong) with the wiring and assembly. Were the assembly instructions wrong? Did they tell you to solder a wire here when it should have been put there? Or were the instructions correct and you, the assembler, slipped up? Or is it something else? Maybe Gizmo, given its design, is supposed to behave this way. Then, of course, the explanation for its behaviour lies with neither a badly printed instruction manual nor incompetent assembly, but with an incompetent, or perhaps just whimsical, designer.

Suppose that, in assembling Gizmo, you mistakenly soldered the blue wire to terminal #16 instead of #18. That is why Gizmo smokes and hums, why it behaves in that strange way, when you turn it on. This, obviously, is a structuring explanation of Gizmo's behaviour. In explaining its behaviour we are citing facts about its internal constitution (the blue wire connected to terminal #16 instead of #18), or what brought that arrangement (your mistake in assembling it), that explain why Gizmo acts the way it does under certain stimulus conditions (when its 'on' button is pushed).

We can, to be sure, shift explanatory interests by changing context. Suppose, after experimenting a few minutes, you turn Gizmo off and set about writing a complaint to the manufacturer. You smell smoke, turn around, and find Gizmo behaving in its accustomed way. Why, you ask, is Gizmo humming and smoking now? Your question now gets a different answer. Gizmo is behaving this way because someone turned it on when you were looking away.

Legitimate though this second explanation is in certain contexts, I am not now interested in it. I am interested in the first answer because of what it reveals about the possibility of giving causal explanations of behaviour that take note of relational facts – historical facts – about how a system came to be in the condition it now is. In the case of some animal behaviour it is, I suggest, structuring explanations of this sort in which information plays a prominent, indeed indispensable, role. It is in the production of such behaviour that animals qualify as semantic engines. It is in the control of such behaviour that it becomes appropriate to speak of the control system as information-driven and, thus, as a mind. To illustrate this quickly, I turn to an example of animal learning.

Suppose an animal – call it Buster – is so wired that it can see nearby *O*s (example from Dretske 1993). Buster not only sees them, it can discriminate them from most other objects. They look different than most other things. This is merely a way of saying that Buster comes equipped with sensory equipment able to deliver information about the presence or absence of nearby *O*s. Buster does not yet know what *O*s are. That will come later. Yet Buster can see them, and seeing them is an informational process: when Buster's eyes are open, something in its brain registers the presence or absence of nearby *O*s just as, when the ignition switch is turned on, a functional gasoline gauge registers the presence or absence of fuel in the tank. Buster's informational relation to *O*s is the same as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

that of human children (with unimpaired eyesight) to objects they see when they have not yet learned what they are.

Because *O*s are dangerous to animals like Buster, it quickly learns to avoid them. Learning to avoid *O*s is a process in which an internal sign of *O*, an internal signal carrying the information that an *O* is present, is made into a cause (a triggering cause) of whatever movements constitute avoidance. The only way to coordinate behaviour (in this case avoidance) with the external conditions in which such behaviour is appropriate (in this case the presence of *O*s) is to make the internal sign of *O*, the perceptual event or events that carry this information, a cause of avoidance. So if Buster learns to avoid *O*s, we know that something inside Buster that indicates the presence of *O*s, something that carries the information that there is an *O* nearby, has been harnessed to effector mechanisms so as to produce output appropriate to the presence of *O*.

From the point of view of understanding how information can play a causal role in determining behaviour, this transformation is significant. As a result of learning of the sort just described, Buster's internal circuitry has been reconfigured so as to give an information-bearing element a control function. An internal *O* indicator assumes causal duties precisely because it carries essential information, information needed to coordinate output with the kind of external conditions in which that output will have beneficial consequences. It is the fact that this element carries information – the information that there is an *O* present – that explains why Buster's control circuits were re-structured so as to produce avoidance when an *O* appeared. The fact that this perceptual event carries this information, therefore, is a structuring cause of Buster's avoidance behaviour. The fact that some internal element carried this information is what explains, causally, why it was made to cause avoidance behaviour. It is, therefore, part of the explanation, the causal explanation, of why Buster avoids the *O*s it sees. The fact that there is something in Buster's brain that carries information about *O*s is as much a part of the causal explanation of Buster's behaviour as the fact that I soldered the blue wire to terminal #16 is a part of the causal explanation of Gizmo's bizarre behaviour.

The conclusion, if not all the steps I took in reaching it, is, I hope, clear enough. If a semantic engine is a system whose performance, some of it anyway, is explained, not simply by the physical events occurring in it, but by the meaning or information that these events carry, then some systems, those capable of the kind of learning just described, are semantic engines. The control system in such engines is an information-driven control system. Such systems are, in this sense and to this extent, minded.

References

Dennett, D. 1983 Intentional systems in cognitive ethology: the panglossian paradigm defended. *Behavl Brain Sci.* **6**, 343–390.

Dretske, F. 1981 *Knowledge and the flow of information.* Cambridge, Mass.: MIT Press.

Dretske, F. 1988 *Explaining behavior.* Cambridge, Mass.: MIT Press.

Dretske, F. 1992 What isn't wrong with folk psychology. *Metaphilosophy* **23**, 1–13.

Dretske, F. 1993 Mental events as structuring causes of behavior. In *Mental causation* (ed. A. Mele & J. Heil), pp. 121–136. Oxford University Press.

Grice, P. 1957 Meaning. *Phil. Rev.* **66**, 377–388.

McGinn, C. 1989 *Mental content.* Oxford: Basil Blackwell.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Discussion

A. Clark (*University of Sussex, U.K. and Washington University, St. Louis, U.S.A.*). What is the role of learning in your account?

F. Dretske. It is the only way in which meaning gets its hands on the steering wheel. I do not see how meaning can have a causal influence in any other way.

A. Clark. Does it matter whether that learning is done by the individual or by evolution? Suppose I was born in a very advanced state, say the state I am in now. I would not have done any of the learning for myself, but surely I would have mental states with meaning?

F. Dretske. Innate behaviours, reflexes and the like, are not mental even when they lead a creature to behave in an intelligent way. When we are discussing minds, the learning must occur during the creature's lifetime. Without the learning, your internal information states do not have the relevant mental content.

Y. Wilks (*University of Sheffield, U.K.*). A statistically based machine translation project in the USA has failed. Should it have worked?

F. Dretske. I do not think my view has implications for technology. My concern is that I do not understand how meanings could be causal. This does not affect much ongoing AI or robotics research.

D. Dennett (*Tufts University, MA, U.S.A.*). What philosophers say is, 'Sure, we know solving that problem is possible in practice, but is it possible in principle?'

R. Chrisley (*University of Sussex, U.K.*). Earlier, you focused on the explanatory role of information and other intentional notions, but later, you suggested this amounted to finding a causal role for informational states. Must informational states be causal, to be explanatory?

F. Dretske. When we give an explanation with a 'because' in it, we must ask what 'because' means, if it is not causal. If it is not causal explanation, then I am not interested. If my thinking is not causal, then why bother thinking at all?

L. J. Cohen (*University of Oxford, U.K.*). Probability is a very awkward problem and there are several different accounts. Which interpretation do you have in mind, and why?

F. Dretske. Propensity theory. (I could not accept a subjectivist interpretation, given my claims that the information is objectively there in the world.) I shall have to live with its unmeasurability. I am only interested in those dependency relations with a propensity measure (or probability) of 1. Certain dependencies exist in the world, such as those brought about by gravity. If probability cannot explain them, then forget probability. It is the dependencies I am interested in.

D. Dennett. I distinguish between semantic and syntactic engines. There are no semantic engines, only syntactic ones that can approximate as close as you like to the performances of a semantic engine. Suppose a person is running an oil refinery. While you are visiting the refinery, the person looks at the dials and yells, 'Run for your life!'. Now, if we replace the person with a machine, and it says 'Run for your life!', do you ignore that because it lacks meaning?

F. Dretske. I do not think fire alarms are semantic engines, but I still run if I hear one sound.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES