Journal for the Philosophy of Science, 34(1983, 97–104.

WHAT IS COMPUTATIONAL PSYCHOLOGY?

Margaret A. Boden and D. H. Mellor

II—D. H. Mellor

I

Professor Boden says that 'computation in the broadest sense is the drawing of inferences from representations' (p. 28). So it is, and I propose to reply by drawing, from this representation of it, some inferences about representations, computations, and how far computational psychology can go.

First, I suppose an inference drawn from one representation is another representation. Next, I take the paradigm of inference-drawing to be a mental activity: the creation of new conscious beliefs from old ones. So representations include conscious beliefs. Suppose I see a rainbow outside and infer that it's raining. My conscious belief that there's a rainbow is (*inter alia*) a mental representation of that state of affairs. My inference creates a new mental representation, of another state of affairs, in the form of my conscious belief that it's raining.

Not all mental representations are conscious beliefs. For a start, not all beliefs are conscious; nor are all inferences. I open a tap to get water because I believe that water only flows through open taps, but I don't bring that belief to mind every time I open a tap. The belief makes me open the tap whether or not it's a conscious belief. Consciously or not, it disposes me to compute representations of an open tap from representations of water flowing; and that disposition itself represents the generalisation that water only flows through open taps.

Nor do conscious and unconscious beliefs exhaust our mental representations. We have other propositional attitudes. I may desire, fear, hope, be glad (etc.) that it's raining. I may contemplate that possibility while wondering what I'll do, or how I'll feel, if it's raining—or if it isn't. All these different states of mind *inter alia* represent the same state of affairs, i.e. that it's raining.

So far mental representations resemble propositions, the classical contents of propositional attitudes. But propositions classically correspond one-to-one to conditions or states of affairs whose obtaining makes propositions true (e.g. the sets of possible worlds in which the propositions are true). Classically, this gives propositions whose truth-values cannot differ, like P and ∼∼∼∼P, the same truth-conditions and thus makes them one proposition. But this makes the proposition that it's raining the same as the proposition that it isn't that it's not the case that it isn't not raining. Yet I can consciously believe the former without consciously believing the latter, as indeed I did until I inferred it by a slow computation on my fingers. They may be the same proposition, but not the same representation: not if representations are what distinguish propositional attitudes of the same kind (e.g. conscious beliefs) held by a single person at a single time.

Maybe they *should* be the same representation. For first, to believe, desire, fear, etc. that P is to believe, desire, fear, etc. that P is true. To have any propositional attitude to any P is to have it to P's truth. To that extent every propositional attitude is concerned with truth. And belief is concerned with nothing else. Alone among attitudes, it aims exclusively at truth: to think something true is always to believe it, but not always to desire, fear, hope or be glad of it. Moore's (1942, pp. 542–3) paradox, the essential absurdity of any assertion of the form 'P [is true] but I don't believe it', has no analogue for other attitudes.

So when I believe P and Q share truth-conditions, I am disposed to believe P if and only if I believe Q. And my other attitudes to P and Q likewise coalesce, despite their not aiming exclusively at truth, so long as I don't think P and Q are necessary truths or falsehoods. (If I do, my attitudes to P and Q may well remain distinct: I can regret that there's no greatest prime number without regretting that 2+2=4.) But provided I think P and Q share contingent truth-conditions, then no matter why I desire, fear, hope, etc. that P (i.e. that P is true), I believe that, necessarily, I will get what I desire, fear, hope, etc. if and only if Q is true. Moreover, I believe I will come to believe or disbelieve that I get what I desire (etc.) if and only if I come to believe or disbelieve Q. How, believing all this, can I desire, fear, hope, be glad, etc. that P without having the very same attitudes to Q?

My belief that P and Q share contingent truth-conditions in effect fuses P and Q into a single content for all my attitudes. If I were an ideal computer, able instantly to compute all identities of truth-conditions (like that of P and ~~~~P), I should always know when my attitudes had the same contingent truth-conditions, and those attitudes would then coalesce. Their contents would be given by their truth-conditions.

But we aren't ideal computers. We compute truth-conditions slowly, fallibly and incompletely. Our attitudes' contents, our mental representations, aren't classical propositions even when we think them contingent, and they certainly aren't when we don't. We may still take them to *represent* the states of affairs, truth-conditions, or sets of possible worlds in which they're true. Only we must then allow many representations of the same state of affairs, to allow us to have pairs of distinct beliefs, desires, fears, etc. that have the same truth-conditions without our realising it.

II

Mental representations differ from propositions in another way, as linguistic representations (sentences) do. 'Representation', like 'sentence' but unlike 'proposition', is used of tokens as well as types. My inferring that it's raining creates a new token representation of that state of affairs, but not a new proposition. Nor a new representation in the type sense in which many people before me have believed, desired, feared, hoped, etc. that it's raining.

The token mental representation here is the embodiment of a propositional attitude in someone during a period of time in which it doesn't change. But it's disputed how attitudes are embodied and therefore what their tokens are. I follow Strawson (1959, ch. 3) in taking attitudes to be properties of whole people: we, not our minds or brains (or their parts), are what represent the conditions in which the contents of our attitudes would be true. Other philosophers disagree, e.g. those who identify mental states with states of the brain. The dispute matters here because what computations are depends on what the token representations are that they deliver and operate on. So I must say why I think people are the tokens at least of their own propositional attitudes.

First of all I should say that, like Fodor (1981, ch. 4), I believe in propositional attitudes. I mean my ascriptions of them to be true, and I think they aren't made true by anything else: not e.g. by people's biology. I take psychology to be as irreducible to biology as biology is to physics (Fodor 1974; Dupré 1981), and for much the same reasons. Facts about what people believe, desire, fear and hope don't reduce to facts about their neurophysiology, let alone to facts about physics.

Moreover, *pace* Davidson (1970), our token attitudes and their changes can have physical causes and effects without being physical particulars. For they needn't be particulars at all: facts, not particulars, are the primary causes and effects. That is, the basic form of causal statements isn't '*d* causes *e*', where *d* and *e* are particulars, but 'Q because P', where P and Q are sentences. 'Q because P' is of course true only when P and Q are (which is why I call causes and effects facts); and as a Humean I suppose 'Q because P' states a causal fact—as opposed say to a proof of Q from P—only when P and Q are logically independent and thus contingent. But causation doesn't relate all contingent facts: a causal 'Q because P' isn't always true when P and Q are. In particular, it won't be true unless Q is more probable than in the causal circumstances it would have been had P been false (Mellor 1985). 'Q because P' is thus only a partial truth-function of P and Q.

(Davidson notoriously argues (1967, pp. 152-3) that if any connective gave the form of causal statements it would have to be completely truth-functional. But his argument fails because it requires the contexts in a causal '. . . because . . .' to be referentially transparent; and they aren't, as inserting suitable identity statements shows. E.g. 'The second take was the best (Q) because (P) the first was only the rehearsal' would if transparent entail both 'The second take was the second take because P' and 'Q because the rehearsal was the rehearsal', and it demonstrably entails neither.)

There are transparent truths of the form '*d* causes *e*', but they are special cases, made true by some true 'Q because P' where P and Q say that descriptions 'F' and 'G' are satisfied, and *d* and *e* are what satisfy them. Basically causes and effects are facts, and so therefore are causal relations. They don't differ from what they relate as universals differ from particulars. If true, a causal 'Q because P', like P and Q, states a contingent fact, which may itself have causes and effects. (E.g. 'He died because he played squash, because he had a weak heart'.) Causation therefore cannot be any one physical relation, nor the province of any one science such as physics, since it relates facts of all kinds, including causal ones. In particular, when a mental fact (P) causes a physical one (Q), or *vice versa*, the causal fact (Q because P) will be both physical and mental.

So my token propositional attitudes needn't be physical in order to have physical causes and effects. Nor need they be particulars distinct from each other (and thus from me) in order to interact. That I have one such attitude needs only to be a fact independent of my having the others, and independence doesn't prevent these facts being about the same particular, i.e. me. When I believe it's raining because I believe there's a rainbow, those are two independent facts—neither entails the other—and both are about me. It is I who by believing there's a rainbow and that it's raining represent those states of affairs. *I* am the token of those beliefs, and of all my other propositional attitudes. They aren't distinct particulars that need identifying with either mental or physical parts of me.

As for token attitudes, so for their changes. I suddenly see that there's a rainbow and so come to believe that it's raining. My successive acquirings of these two beliefs are changes in me. But not changes *within* me, i.e. not changes in parts of me. Naturally they don't occur by magic, unaffected by what happens to my mental and physical parts. I only come to believe there's a rainbow because photons enter my eyes, and because resulting changes within me then make me behave as people do who believe there's a rainbow—e.g. by making me infer that it's raining. But none of these changes within me is my change of belief, any more than the opening of a car's carburettor, or the consequent angular acceleration of its wheels, is identical with the acceleration of the car as a whole.

Still, many of these changes in and within me, that make me believe it's raining, could be psychological, and could be computations. The last one certainly is. When I infer rain from a rainbow I undoubtedly compute one representation from another. But the first one isn't. The change that makes me believe there's a rainbow starts from photons entering my eyes; and though the photons come from the rainbow, their arrival certainly doesn't represent it.

But the incoming photons do other things to me before making me believe there's a rainbow. Seeing is a many-stage process, whose later stages may be computations even if the first is not. And according to Marr (1982), Boden's paradigm of computational psychology, they are. The process of seeing starts, he says, with 'arrays of image intensity values, as detected by the photoreceptors in the retina' and proceeds 'by mapping from one representation to another' (p. 31) until it delivers a belief about 'what is present in the world and where it is' (p. 3).

But what makes my photoreceptors' response to an array of image intensity values a representation of it? The response betokens no belief of mine about the array (and if not a belief, certainly no other propositional attitude). For first, I, not my retinas, should be the token of that belief. My retinal responses could at most be its mechanism, as a brake pedal's depression is of a car's stopping. But anyway I have no beliefs, not even unconscious ones, about arrays of image intensity values across my retinas. Maybe I could have now that reading Marr has given me the necessary concepts. But not before, and though reading Marr has undoubtedly improved my mind, it hasn't improved my eyesight.

And as for this stage in the process of seeing, so for the others. To see a rainbow I need have no beliefs, and no other attitudes, whose contents are the representations Marr says occur *en route* to my coming to believe there's a rainbow. Two possibilities remain. Either parts of my optic nerves have the beliefs I lack, about arrays of image intensity values, etc. Or these token representations aren't token attitudes at all. In the next two sections I consider these possibilities in turn.

III

Do my retinas, or other parts of my optic nerves, have beliefs? I think not, though I shall give the thesis the benefit of every reasonable doubt. I shan't for instance object that we don't normally credit retinas with beliefs: we don't understand the mind well enough for that to prove much. We may understand it well enough to rule out the discovery that we have no beliefs; but not the discovery of beliefs within us that we didn't know we had. That possibility follows from the discovery of unconscious beliefs which I've already granted. So my optic nerves needn't lack beliefs because they lack consciousness. Nor because they lack language: some languageless animals I suppose have beliefs. Languageless believers are admittedly more debatable than unconscious beliefs (e.g. Davidson 1975), but that doesn't matter here. There are better reasons for denying optic nerves beliefs than their inability to talk.

Nor need my optic nerves being parts of a believer deprive them of beliefs. They could be tokens of their own beliefs without being tokens of mine. And believers can get beliefs and other attitudes from those of their parts. 'Social' believers such as churches, firms and unions do just that (Mellor 1982), as does Boden's 'class of highly opinionated schoolchildren' [p. 30-1]. It has desires—those of its noisiest members—and believes whatever its odd 'opinion-forming process' would lead them to agree on.

But not all believers are social in that sense: the regress will stop somewhere. In fact (*pace* Dennett 1978), it stops with us. We don't get our propositional attitudes from those of our parts, because they have none. And they have none because their states don't combine to cause their activities as our attitudes do to cause ours—especially our beliefs and desires. For what makes these states of ours beliefs and desires *is* the way they combine to make us act. Our consciousness of them won't characterise them (Mellor 1980), but the way they combine to cause our actions will, at least in part. Suppose for instance I make P true because I desire Q and believe making P true will make Q true. My action is caused by a combination of belief and desire; and so are all actions. What my beliefs make me do always depends on my desires and *vice versa*; and neither belief nor desire on its own would make me act at all. But that's not how the parts of my optic nerves act. Their 'beliefs' about arrays of image intensity values and the rest needn't combine with anything to cause their output: the 'belief' *is* the output. Calling it a belief is either vacuous or false—either adding nothing or imputing a structure to its causation that it doesn't have.

In short, our optic nerves and other parts lack beliefs because they lack desires; and lack both because the states that cause their activities don't combine to do so as beliefs and desires do. And if they lack beliefs and desires, they certainly have no other attitudes.

As the above argument indicates, I think propositional attitudes must be functionally defined (cf. Block 1980, Pt. 3). That is, I take them to need fixing by their perceptual and other stock causes, by their interactions, and by their behavioural effects. My optic nerves have no attitudes because their states don't satisfy such conditions. But that doesn't rule out computational psychology. It only prevents my optic nerves' responses being token attitudes. They may still be token representations of some other kind, as we shall see in section IV.

Meanwhile, computational psychology, far from competing with functionalism, both needs and is an instance of it. It needs it because computations only transform representations, which are the contents of all kinds of attitudes. Believing that P, desiring that P and fearing that P are different mental states, but they all have the same content. As representations, i.e. as sources or upshots of computations, they are all the same. Computational psychology on its own won't make them different. Functionalism will. My belief that P differs from my desire that P because, as functionalism says, different things induce these states in me and they affect my actions differently.

Computational psychology is an instance of functionalism because it identifies representations by their interactions and other causes and effects. It makes believing there's a rainbow differ from believing it's raining by the different causes, e.g. arrays of image intensity values, of those two beliefs; by our computing the latter from the former (but not *vice versa*) and the belief that if we go out we'll get wet from the latter (but not the former); and so on. The inferential, i.e. computational, role of a mental state is clearly part, though not all, of its functional role.

Functionalism incidentally fits Boden's idea of people as 'connectionist systems . . . made up of locally communicating units functioning in parallel, where—because of excitatory and inhibitory connections—the state of any one unit depends largely on the states of its neighbours' [p. 30]. The fact that I believe it's raining is just such a unit, exciting some and inhibiting other facts about my beliefs, desires, fears, hopes, etc.: i.e. about other attitudes of mine with which that belief communicates and functions in parallel, and on which its content, like each of theirs, largely depends.

But functionalism is a disputed doctrine, so I should argue for as well as from it, though I can't do that properly here. Still, I can meet some objections to it, which really apply only to the physicalism that many functionalists have needlessly espoused. Take the objection that functionalism can't cope with 'qualia'—sensations, pains, feelings, etc. (Fodor 1981, p. 16). Free of physicalism, functionalism needn't cope with qualia at all. Qualia don't need defining: they are self-individuating states of consciousness, in terms of which other states—allergies, susceptibilities to light, sound and other sources of sensations and feelings—can themselves be functionally defined.

Functionalism also copes better on its own with other kinds of consciousness. 'First-order' attitudes, defined by the bodily behaviour they combine to cause, can be used to define 'higher order' attitudes that account for our consciousness of first-order ones. Becoming conscious of a first-order belief, for instance, is just coming to believe one has it (Mellor 1980). These functionally defined conscious states enable us in turn to define computational and other abilities by the role they play in our conscious thought. Functionalism on its own can exploit and explain our inner mental life as easily as our bodily behaviour.

What it can't so easily explain are the outward contents of our propositional attitudes. Suppose my belief that it's raining makes me take my umbrella, because I want to keep dry. That belief is partly defined by that effect. But the belief may be false. I may have it when it isn't raining: I will still take the umbrella. The conditions in which my beliefs—and other attitudes—make me do something aren't those in which their contents are true. Their behavioural effects won't fix their truth-conditions. Nor will their internal interactions, which will also be the same whether their contents are true or false. Not even their external causes will give attitudes their truth-conditions. Rain obviously won't be what makes me desire rain, or fear it or hope for it. It may make me believe in it, but even that won't help. Our beliefs' causes can't be their truth-conditions since (a) they have too many causes, and (b) we must be able to mistake what we see. Functionalism can use the photons that make me believe it's raining to help define the content of that belief, but not the rain.

Now this doesn't show that functionalism is wrong, merely that, like patriotism, it isn't enough. For the rest I would look to the conditions in which our actions succeed, i.e. which give us what we acted to get. That happens when, though not only when, all the beliefs we act on are true. The conditions in which actions succeed include the truth-conditions of the beliefs that cause them (cf. Putnam 1978, p. 99). That fact—the grain of truth in the pragmatic equation of true beliefs with useful ones—should at least help to fix the truth-conditions of beliefs.

But even as it is, functionalism will do for a 'narrow' psychology that leaves others to give mental representations their truth-conditions (Fodor 1980). And all agree that a 'wide' psychology won't be computational unless a narrow one is. But is it? Functionalism already prevents my optic nerves' responses being token attitudes. How then can they be representations?

IV

How can a retina's response to an array of image intensity values represent it without being a token attitude towards it? First, by changing whenever the array changes in whatever respect it represents. But there must be more to it than that. For most things expand when heated: their volumes change when their temperatures do. Yet not all those volumes represent temperatures. Why not, if retinal responses represent retinal arrays?

Perhaps retinal responses are representations because they set off computations, i.e. because some of their effects are computed —which makes the effects representations too. Whereas the effects of volumes generally aren't computed. A bubble doesn't compute its surface area (or anything else) from its volume, so neither its volume nor its area represents its temperature or anything else.

But why suppose retinal responses fix their effects by computation? Not because we do. We can compute a bubble's area from its volume, i.e. infer a belief about the area from a belief about the volume. No one infers from this that the area itself is computed from the volume (cf. Dennett 1977).

The inference is more tempting, though no more valid, when what fixes something is itself a representation. Suppose I use decision theory to infer what you will do from what (I believe) you believe and desire. That is, I compute a representation of your action from representations of your beliefs and desires, which are also representations. Suppose even that you do this, forming beliefs about your own beliefs and desires by bringing them to consciousness in order to predict your own action. It still doesn't follow that the action itself is computed from your beliefs and desires, and we shall see in section V that it isn't.

It likewise doesn't follow, from our inferring each stage in the process of seeing a rainbow from the stage before, that the stages themselves are produced by computations, even if they are representations. We know the last one—the belief that there's a rainbow—is, but that doesn't mean its precursors are. Not even Marr thinks a retinal response's precursor—the array he thinks it represents—itself represents anything, or that the response is computed from it.

What then *does* make retinal responses represent retinal arrays, and the later stages of seeing computations? Consider the paradigm of token representations that aren't token attitudes: written or printed token sentences. Perhaps retinal responses are like token sentences. How do tokens of 'It's 20°C' manage to represent that state of affairs when the volumes of bubbles at 20°C don't?

Token sentences represent what we use them to represent, and our use is conventional. That is, *we* decide arbitrarily what tokens of 'It's 20°C' shall represent: that array of symbols isn't naturally linked to a temperature as the volume of a bubble is. But retinal responses aren't conventional. The makeup of the eye links them as naturally to the arrays that induce them as bubble volumes are linked to temperatures. If representations have to be as conventional as linguistic ones are, retinal responses will certainly not qualify.

But they may still satisfy weaker conditions that token sentences satisfy and bubbles don't. Boden's (1980) discussion of cognitive biology suggests one: that representation should be, if not conventional, at least arbitrary. That is, what a token represents shouldn't be fixed by what it's made of. (Just as the ink used to print tokens of 'It's 20°C' isn't what makes it represent a thermal state of affairs.) What it represents should depend on context, as tokens of 'It's 20°C' (i.e. 20°C here and now) do on where and when they're produced. These are the features that Boden (1980, pp. 42-5) takes to excuse Goodwin's (1976) describing certain biological processes in cognitive terms.

Unfortunately they are also features of bubbles. The temperature needed to give a bubble a certain volume isn't fixed by what it's made of, i.e. by the gas in it. It depends on how much gas is in it, and that isn't fixed by what gas it is. It also depends on the context, i.e. on the external pressure. In short bubbles respond just as arbitrarily (and just as well) to their temperatures as retinas do to their arrays of image intensity values. If biology is cognitive by these criteria, so is physics. But when all grass is made flesh, vegetarians become carnivores and the whole point of their protest disappears. And making psychology computational by standards that make biology and physics so tells us nothing special about psychology.

We must do better than this for computational psychology, and we can. Consider why we use sentences to represent states of affairs. We use them to communicate our propositional attitudes, especially our beliefs. The object of linguistic conventions is to make tokens of specific sentence types induce specific attitudes in the people we expose to them. And attitudes are representations. Token sentences become representations by being used to induce token attitudes, and get their contents from those of the attitudes they induce.

(It is of course disputed whether sentence types get their meanings from the contents of the attitudes we use their tokens to communicate, or *vice versa.* Is the English meaning of 'It's raining' given by the content of my belief that it is or the other way round? Here fortunately it doesn't matter. Either way our using that sentence's tokens to induce belief, desire, fear, hope, etc. that it's raining is what makes them represent that state of affairs.)

As for token sentences, so incidentally for some token volumes. The volume of working fluid in a thermometer *does* represent its temperature (and hence that of whatever it's used to measure), because that's what we make it induce beliefs about. Fluid volumes generally don't represent temperatures, merely because they aren't generally used to generate beliefs about them.

Thermometry of course depends on conventions as much as our use of sentences does, e.g. on conventional temperature scales. But not all representations are as conventional as that. Photographs aren't, nor are films, plays, paintings and other non-linguistic representations of states of affairs. No doubt most of them use some conventions, but not as many as language does.

Photographs represent landscapes much less conventionally than descriptions of them do; and though the ideal landscapes 'Capability' Brown's creations represent may be conventional, his representation of them needs no conventions, for his landscapes represent themselves.

Representations need to be, not conventional, but made—or evolved—to induce attitudes, whose contents then fix theirs. Which is just what the earlier stages of seeing are. They have evolved to induce specific beliefs, just as photographs are made to do. If photographs can therefore be representations, why not these internal precursors of beliefs?

It is no objection that the early stages of seeing don't represent what the belief they induce represents. Nor need a photograph. Photographs represent what things look like from a certain place. But the beliefs they are apt to induce needn't represent what the things in the photograph look like. (Consider a still photograph of someone running.) They need only represent something that such beliefs will be inferred from. And so it is with the stages of seeing. My seeing, if Marr is right, is a sequence of inferences from 'viewer-centered' representations supplied by my retinas to my 'object-centered' beliefs about rainbows and the like (Marr 1980, p. 37).

The representations in this sequence must then be such as to yield by these computations the belief I get at the end of it. And the first representation has to be set off by photons entering the eyes. The contents of the stages in between are fixed by these two constraints. Each computation must be an actual response to the stage it starts from (as a thermometer's display is to the fluid volume that represents its temperature). And the sequence of computations must eventually transform our retinas' initial representations of arrays of image intensity values into the contents of our beliefs about what we see. And finally, the whole process must be as reliable as we know our vision is. That is, it must turn arrays of image intensity values into beliefs that generally represent, i.e. are true in, the very conditions which produce the arrays that yield them.

If we see like this, then we see by computation. The thesis that we do is neither trivial nor incoherent. Nor is it trivial to make a testable theory of vision out of it as Marr has done by specifying the stages and computations involved. Right or wrong, his theory shows what computational psychology can do. But could we compute a summer from this swallow? How much further can computational psychology go?

V

Computational psychology, on Boden's 'minimal definition', is 'the study of the various computational processes whereby mental representations are constructed, organised and transformed' [p. 17]. Marr shows how seeing may be one such process. But there aren't many others. Computation in Boden's 'broadest sense . . . the drawing of inferences from representations' [p. 28] won't account for much of the mind.

First, many mental states, such as sensations and pains, are not representations. They may be self-intimating, i.e. make us believe we have them, and from those representations of them we may compute others, i.e. beliefs about their causes. But that doesn't make the states themselves representations, nor their transactions computations. They can't be, since they have no true or false contents to be the premises or conclusions of inferences. Nothing about them is capable of truth or falsity: they have no truth-conditions and therefore represent no states of affairs.

Nor do they get truth-conditions from the beliefs they cause, as my optic nerves' responses do from those they cause. Perceptual beliefs don't give truth-conditions to what they are about, which is what they would need to do here. My seeing that there's a rainbow doesn't make the fact that there's a rainbow represent anything. If it did, it could only make that fact represent itself, which won't tell us what fact it is. So even if my being in pain represents itself because it makes me believe I'm in pain, we must still say what pain is to give that representation's truth-conditions; and similarly for sensations.

But in the sense of section III, computational psychology is 'narrow': it leaves others to give its representations' truth-conditions. Computational psychology won't say what a belief that it's raining represents, i.e. what rain is; nor likewise what a belief about a pain or a sensation represents, i.e. what they are. But whatever they are, they're mental. Saying what they are, what sorts we have and why, what causes them and what they cause, is psychology of some kind. But not computational psychology. *We* may compute answers to these questions, but the answers won't refer to computations.

Computational psychology's restriction to representations confines it to the contents of propositional attitudes and of some of their mental causes. That is, to what belief, desire, fear, hope, etc. that P have in common. But since tokens of these attitudes all have the same content, computation won't explain how they differ. As we saw in section III, computational psychology on its own must treat them all alike, as identical representations (=P). Yet obviously they differ, even computationally. I may compute a belief that P from a belief that P & Q but not a fear that P from a fear that P & Q (I may only fear Q). So computational psychology must let computations to and from a token attitude to P depend on what the attitude is. But then it needs some other account of attitudes. Our theory of mental computation will have to be based on our theory of propositional attitudes, not the other way round.

As for difference, so for change. When my fear that P changes into a belief that P, there is no change in the representation, P. The change isn't a computation, and computational psychology won't explain it. For as we've just seen, it won't explain the difference between fearing P and believing P that makes this the change it is. Boden's doubts about the prospects for a computational explanation of attitude-change [p. 28] are well founded.

For the same reason computational psychology won't cover the psychology of action. Suppose I do P (make P true) because I desire Q and believe that Q will be true if and only if P is. It takes more than a computation from Q and Q↔ P to make me act: I shouldn't do P if I feared Q instead of desiring it. Our actions depend not just on the contents of our attitudes but on which attitudes have which contents, which isn't a computational matter at all. My action may be *predicted* by computing a belief that I will do P from a belief that I desire Q and believe Q↔ P; but as we saw in section IV, so may the volumes and areas of bubbles be predicted by computation. In neither case are the facts that verify the predictions caused by computations.

Computational psychology can at most cover processes that 'construct, organise or transform' the contents of a single propositional attitude, especially belief. Inferring one belief from another is the paradigm of computation. But it isn't something that computational psychology explains. On the contrary, it's what gives the theory of computation, i.e. of inference, its sense. Rules of inference, for instance, should preserve truth; so computational processes should be 'reliable', i.e. compute false representations from true ones as rarely as possible. But this is only a virtue because, as we saw in section I, belief aims solely at truth. When representations aren't beliefs, reliability may not be a virtue at all. Suppose I fear that Q because I fear that P. There's no virtue in that computation being reliable, i.e. in Q being more probably true if P is than if it isn't. On the contrary, I shall do my best to make Q false whether P is true or not.

Calling computing 'information-processing' (Boden, section II) also presumes the virtue of preserving truth. The concept of information derives entirely from belief's peculiar and exclusive concern with truth, and applies to no other attitude. Believing P may embody the information that P: desiring, fearing or hoping that P doesn't. (Being glad that P does, but only because it entails believing P.) Whatever computing one desire, fear or hope from another is, it isn't processing information.

The fact is that computational psychology requires an ideal of truth-preserving inference which makes sense only of computations leading to beliefs. That indeed covers perception, whose function is to produce true beliefs about the world. My retinal responses to arrays of image intensity values can be belief-like representations of them. That is, they can aim only to be true, i.e. to be such as to lead, if the subsequent computations are correctly carried out, to true beliefs about what I see. Computational theories of seeing and other ways of perceiving make perfect sense and may well be true.

Computations leading to attitudes other than belief couldn't be governed just by the ideal of preserving truth: truth isn't their sole object, and sometimes, as with fear, isn't their object at all. But computational theories have no other ideals to offer. The contents of propositional attitudes represent nothing but truth-conditions: they have nothing but truth-values for computations to affect. So computational psychology not only can't say how other attitudes differ from belief and from each other: it can't even say how or why their contents change. That is to say, they don't change by computation.

In short, the case for computational psychology is a fallacious extrapolation from the one part of psychology—the aetiology of belief—that can be computational. As the whole subject, computational psychology is, as Boden says, a dragon. We don't know much about it; but we know enough to know there's no such thing.

NOTE

I am especially indebted for ideas expressed in this paper to work of Professors Strawson, Dennett, Fodor and Anthony Appiah, and material in Woodfield (1982); to discussions at the September 1983 meeting of the Thyssen UK Philosophy Group, and at a Cambridge research seminar on functionalism given with Anthony Appiah in the Michaelmas Term 1983; and to conversations with Jeremy Butterfield, David Papineau, Diane Shard and Timothy Smiley.

REFERENCES

N. Block, ed. (1980), *Readings in Philosophy of Psychology: Volume I*, Methuen.

M. Boden (1980), The case for a cognitive biology I. *Aristotelian Society Supplementary Volume* 54, 25-49.

D. Davidson (1967), Causal relations. *Essays on Actions and Events* (1980), Clarendon Press, pp. 149-62.

D. Davidson (1970), Mental events. *Essays on Actions and Events* (1980), Clarendon Press, pp. 207-25.

D. Davidson (1975), Thought and talk. *Mind and Language*, ed. S. Guttenplan, Oxford University Press, pp. 7-23.

D. C. Dennett (1977), A cure for the common code? *Brainstorms* (1979), Harvester Press, pp. 90-108.

D. C. Dennett (1978), Artificial intelligence as philosophy and as psychology. *Brainstorms* (1979), Harvester Press, pp. 109-26.

J. Dupré (1981), Natural kinds and biological taxa. *Philosophical Review* 90, 66-90.

J. A. Fodor (1974), Special sciences. *Representations* (1981), Harvester Press, pp. 127-45.

J. A. Fodor (1980), Methodological solipsism. *Representations* (1981), Harvester Press, pp. 225-53.

J. A. Fodor (1981), Introduction. *Representations*, Harvester Press, pp. 1-31.

B. C. Goodwin (1976), *Analytical Physiology of Cells and Developing Organisms*, Academic Press.

D. Marr (1982), *Vision*, Freeman.

D. H. Mellor (1980), Consciousness and degrees of belief. *Prospects for Pragmatism*, ed. D. H. Mellor, Cambridge University Press, pp. 139-73.

D. H. Mellor (1982), The reduction of society. *Philosophy* 57, 51-75.

D. H. Mellor (1985), Fixed past, unfixed future. *Contributions to Philosophy: Michael Dummett*, ed. B. Taylor, Nijhoff.

G. E. Moore (1942), A reply to my critics. *The Philosophy of G. E. Moore*, ed. P. A. Schilpp, Northwestern University Press, pp. 553-677.

H. Putnam (1978), *Meaning and the Moral Sciences*, Routledge.

P. F. Strawson (1959), *Individuals*, Methuen.

A. Woodfield, ed. (1982), *Thought and Object*, Clarendon Press.