Chapter Four:

Situated cognitive science and emotion

> *In contemplating your true self, don't include the body which surrounds you and the limbs attached to it. They are like tools, the only difference being that they grow from the body.*
>
> Marcus Aurelius, *The Meditations*

A growing number of cognitive scientists today argue for the need to take a 'situated' approach to the study of the mind. In this chapter, I outline the main principles of this approach, and show how they can be applied to the interruption theory of emotion.

4.1. Situated cognitive science

When the term 'evolutionary' is used to describe an approach to the study of the mind, one immediately gets a pretty good idea what is involved. The same cannot be said of the term 'situated'. The label is due to Lucy Suchman, whose 1987 book, *Plans and Situated Action*, has been one of the most influential works in shaping this approach (Suchman, 1987). The core idea in Suchman's book, as I see it, is that minds are 'leaky', although this way of putting things is due to Andy Clark rather than Suchman herself (Clark, 1997).

The leaky mind

In section 2.1, I identified internalism as one of the assumptions of classical cognitive science. I also argued that the best way to construe this term was as the view that the boundary of the mind is to be *identified* with some physical feature of the organism. This is the assumption that the proponents of the situated approach dispute. The input-output boundary is, according to them, a moveable feast. In other words, the question of where the input-output boundary is to be physically located will depend very much on the particular context of enquiry. The boundary of the *mind* is not to be simply *identified* with *any* physical boundary, whether that between the brain and the rest of the body or that between the body and the rest of the world. Some mental processes may well supervene entirely on the brain, or even on just one part of the brain. Others supervene on the brain plus part of the body. Others supervene on the brain, body, and parts of the world. This is what I mean by saying that minds are 'leaky'.

Why should we take the idea of the leaky mind seriously? Without attempting an exhaustive survey, I will summarise some of the arguments here. Since the most ‘watertight’ view of the mind locates it entirely in the brain, I begin by looking at cases in which the mind leaks out into other parts of the body. I then go on to examine cases in which the mind leaks even further afield, spilling out of the body into the world around it. If we wish to distinguish between these two cases of leakage, we could refer to the first as examples of the ‘embodied’ mind, and the second as cases of the ‘embedded’ mind.

The embodied mind

The idea that the mind can be *embodied* is more than just the claim that natural minds are always found in creatures with bodies. The latter claim is just STM. To say that minds can be *embodied* is to say that mental processes do not always supervene simply on the brain; sometimes, they supervene on the brain *plus some other part of the body*.

One kind of argument for the idea that minds can be embodied in this sense appeals to cases in which changes to the body cause changes to the mind. For some of those afflicted by Möbius syndrome, for example, the inability to move any of the muscles of facial expression leads the person to experience significantly less affect than others (Cole, 1998). In such cases, it is not plausible to think of the body as simply a medium through which the mind expresses itself; the body affects the mind so powerfully that it seems more than just a figure of speech to say that it is part of the mind. Daniel Dennett makes a similar point in his book, *Kinds of Minds*, where he argues that the mind is ‘distributed’ throughout the body. In Dennett's words, 'my body contains as much of *me*, the values and talents and memories and dispositions that make me who I am, as my nervous system does' (Dennett, 1996: 77, emphasis in original)

Those who are tempted to dismiss such cases on the grounds that they do not deal with properly *cognitive* phenomena can be presented with other cases. People often use their fingers to keep a tally of something instead of storing the number in their heads. Less consciously, there is no need for the brain to calculate the force needed to sit up straight too exactly, since a certain amount of leeway is catered for by the purely physical properties of the body. Andy Clark takes these kinds of example to show that other organs besides the brain are integral to the processing loops that result in intelligent action. In other words, the brain can ease its computational burden by ‘offloading’ some information processing onto other parts of the body (Clark, 1997).

More generally, George Lakoff and Mark Johnson claim, in *Philosophy in the Flesh*, that the very structure of our thoughts depends on the particular kind of body we have. For example, ‘the fact that we have muscles and use them to apply force in certain ways leads to the structure of our system of causal concepts' (Lakoff and Johnson, 1999: 19). According to Lakoff and Johnson, if we had different kinds of bodies, we would have different kinds of concepts. This thesis could be tested empirically. We might, for example, be able to design an experiment to test whether people with only one arm could conceive of 'weighing' and 'balancing' as easily as those with two.

The embedded mind

According to the leaky mind hypothesis, the mind is not only capable of spilling out of the brain into the rest of the body, but also of flowing out of the body into the rest of the world. In addition to being embodied, minds can be *embedded*.

One set of examples to support this view involve 'cognitive artefacts' – objects made for the purpose of aiding cognition, such as maps, notepads and computers. Such objects not only lighten the processing load of the thinker (by allowing, for example, external memory storage), but also make possible certain tasks that would otherwise be practically impossible. Nicholas Humphrey puts the point well when he writes that a man with a pair of scissors 'is not just handier, he is in effect brainier – because he can now exploit his brain power in new ways' (Humphrey, 1997: 100).

Daniel Dennett and Andy Clark have both argued that the most important among the various mind tools at our disposal is language. The idea that external linguistic tools might alter and inform an individual's intrinsic mode of information-processing goes back at least to the Soviet psychologist, Lev Vygotsky, but for many years this possibility was largely ignored by many Western cognitive scientists, who followed Chomsky in adopting a thoroughly internalist approach to the study of language. Dennett and Clark have helped to redress this imbalance by showing how 'language is not just a medium *in which* we think', but in fact 'actually *does* some of the thinking with us and for us' (Humphrey, 1997: 101, emphasis in original). In Clark's terms, language is the 'ultimate (cognitive) artefact'; it provides the human brain with its most extensive and powerful 'external scaffolding'. Much of what humans take to be their own mental capacities, argues Clark, rely crucially on the external scaffolding provided by language (Clark, 1997).

The mind can leak out even further into the rest of the world, to include other people as well as cognitive artefacts. Edwin Hutchins proposes that the cockpit of an aeroplane be regarded as a cognitive system in which cognition is *distributed* over the pilot, co-pilot, and navigator, as well as over the various objects they use (Hutchins, 1995). There are, he claims, certain cognitive processes which simply cannot be located in one particular part of this system. Some of the complex capacities that we identify as mind and intellect may be much more like the systemic properties of the cockpit of an aircraft than intrinsic capacities of the bare biological brain.

The internalist reply

The internalist can respond to all the supposed cases of embodied and embedded minds with a single, simple move; he can simply stick to his guns, and insist that what is going on in such cases is simply a case of a brain interacting with its environment. If we want to tell a richer story about how this two-way interaction goes on, all well and good, but the mind still supervenes exclusively on the brain.

The proponent of the situated approach cannot deny this move on purely logical grounds. He must concede that, for all the cases of leakage described above, it is always logically possible to redescribe them in such a way as to shift the input-output boundary back onto the skin of the organism, or further back still, onto the brain-body boundary. His reasons for rejecting such a move must, therefore, be of a less abstract kind than pure logic. This, however, is nothing new. Pierre Duhem and W. V. O. Quine showed long ago that scientific questions can never be decided on logic alone, nor on facts alone. Most philosophers of science today accept that our choice of alternative theories must be guided by other more pragmatic grounds too.

Two such pragmatic criteria present themselves as relevant to resolving the debate between internalism and the leaky mind hypothesis. The first is explanatory flexibility, and the second is empirical fruitfulness.

By explanatory flexibility, I mean that our explanatory tools must be able to accommodate themselves to different kinds of situation, rather than forcing the facts to accommodate to the theory. I recognise that this is a hopelessly sloppy way of putting things, but I take it that the reader has some idea of what I mean. The metaphors of the shoe-horn and the Procrustean bed are often used to describe inflexible explanatory frameworks. Of course, we don't want our framework to be *too* flexible, so that it can accommodate itself to any fact at all, for otherwise it will become vacuous. But nor do we want it to be so inflexible that every single situation receives the same pat response. That seems like dogmatism rather than science.

The internalist response seems to have all the hallmarks of dogmatism. No example of embodied or embedded minds, it seems, will be sufficient to persuade him that the input-output boundary is flexible. Yet he can give no principled reason for taking the skin or any other purely physical thing to be so crucial to the concept of mind. In fact, this seems to fly in the face of the key idea of substrate neutrality, which is so central to cognitive science. The brain is only one element in the processing loops that result in intelligent action, and to single it out as somehow 'more vital' to these action loops than any other seems somewhat arbitrary (Clark, 1997). Richard Dawkins has levelled the same charge of arbitrariness at biologists who identify the phenotype with the physical organism (Dawkins, 1982). Indeed, Dawkins' concept of the 'extended phenotype' can be seen as the biological analogue of the idea of embeddedness.

Parsimony

The second pragmatic criterion that is relevant to resolving the debate between internalism and the leaky mind hypothesis is empirical fruitfulness. By empirical fruitfulness I mean that we should prefer one theory over another when it leads, as a matter of fact, to more interesting technical achievements. In the context of the debate between internalism and the leaky mind hypothesis, these technical achievements are to be found in the domain of artificial intelligence. Thus, if we find that cognitive scientists who adopt the leaky mind hypothesis tend to design better computing machines than those of a more internalist bent, this will constitute good grounds for preferring the leaky mind hypothesis. By their works shall ye know them.

The relevant sense of 'better' in this case is not merely one of performance, but one of accuracy in modelling natural minds. This assumes that we have some independent criterion, other than performance, for judging which of two machines is a more accurate model of the human mind. I think we do have such a criterion; it is the criterion of parsimony.

In this context, the criterion of parsimony states that whenever two machines are capable of mimicking some aspect of human behaviour equally well, we should assume that the simpler machine is a better model of the human mind. The justification for using this criterion is not just that simpler models are easier to work with. Nor is it just that the criterion accords with Occam's razor. It is also that, other things being equal, natural selection will favour simpler minds over more complex ones. To get this extra justification for the criterion of parsimony, of course, we need ETM.

When we compare the machines designed by proponents of the situated approach with internalist-type machines, the criterion of parsimony often favours the former. For example, how might we construct a robot that can pick up disused cans from the office floor? If we took a typical internalist approach we might assume that it was necessary to give the robot a sophisticated internal map of the office, and a complex camera system to register its position. The camera could search for tin cans at long range, in which case we would also need some complex image-analysis software. Alternatively, we could give the computer a short-range can detector and then make it traverse the whole office according to some pre-arranged systematic route. Either way, the program is going to be quite complex. Let's call this the internalist robot.

Connell, however, took a rather different approach. Inspired by the idea of offloading as much computation as possible on the environment, Connell simply allowed his robot to take a random route and gave it a host of infrared proximity sensors to help it navigate along walls and through doorways. Whenever it happened to detect a can with its short-range metal detector, it stopped to pick it up. This robot performed its task perfectly well without the need for any internal map of the room, nor for any pre-arranged route to be stored in its memory. As Rodney Brooks argues, it is natural to construe the robot built by Connell as being far simpler than the internalist robot (Brooks, 1991). This is only one example, but in many other cases too, the machines designed by those adopting a situated approach have managed to achieve comparable results to more internalist-style machines despite being much simpler. Overall, then, when judged by the criterion of parsimony, the situated approach has led to the design of better artificial minds than the internalist approach.

Heuristics

There is nothing about CTM that forces one to adopt an internalist approach. CTM is perfectly compatible with the leaky mind hypothesis. Nevertheless, the fact remains that most classical cognitive *did* adopt an internalist approach. It seems that, unless people are forced to think otherwise by adopting an explicitly situated approach, they tend to operate on internalist assumptions.

Likewise, there is nothing about the internalist approach that *forces* one to design robots in one way rather than another. There is nothing about adopting the internalist stance that obliges the roboticist to design what I have dubbed 'internalist-robots'. Still, those who have adopted an internalist approach have tended to ignore the possibility that computational burdens can be eased by simple mechanical solutions. Even though there is nothing about the internalist position that logically excludes this possibility, it is the case that, as a purely empirical matter, cognitive scientists of an internalist bent have regularly ignored it. The situated approach recommends itself to us then, along practical grounds too. It is a heuristic that can guide those working in artificial intelligence to design their machines in simpler and more elegant ways.

I have argued that the situated approach can shed new light on the nature of mind. In the next section, I argue that the same is true when it comes to understanding that particular subset of the mental we refer to as emotions.

4.2. Situated approaches to emotion

In section 3.2, I argued that new light could be thrown on the emotions by defining them as interruption mechanisms. In this section, I argue that we can advance our understanding of the emotions still further by applying a situated approach to the interruption theory.

The key idea of the situated approach is that mental processes are leaky. They do not always supervene on the brain alone. Sometimes they are 'embodied', in the sense of supervening on the brain plus some other part (s) of the body. And sometimes they are 'embedded', in the sense of supervening on the extra somatic environment too. In this section I argue that this is particularly true of those mental processes we call emotions. In other words, interruption mechanisms are very often embodied and embedded.

The internalist view of emotion

The most obvious roles for bodily organs other than the brain in emotion is to serve as sensors for gathering emotionally relevant data and to serve as media for communicating emotions to other cognitive agents. These roles were clearly recognised by the proponents of the classical approach to emotion, but they tended to focus exclusively on *linguistic* inputs and outputs. The paradigmatic model of such a purely linguistic system is, of course, the Turing test, which (in its classical form) limits all communication to text messages sent via a keyboard to an alphanumeric display. In such a set-up, emotions can only be expressed by linguistic means. However, most emotion researchers today accept that the 'affective bandwidth' of linguistic media is very narrow, notwithstanding the expressive power of written dialogue that we find in great scripts and screenplays. Emotions are primarily communicated not by words but by a whole host of nonlinguistic bodily signals, from facial expression and vocal intonation to gestures and posture. To make our computer models of emotion more realistic, therefore, we should give them the ability to use the visual, auditory and physiological signals of emotion. To this end, roboticists and others working in the field of affective computing are increasingly supplying their machines with a wide variety of peripherals in addition to the standard keyboard and monitor. Peripherals that could be such as cameras and microphones are used to help computers get better at recognising emotions in humans, while animated agent faces and affective voice synthesisers are used to model nonlinguistic forms of emotional expression (Picard, 1997).

All this, however, is perfectly compatible with an internalist approach to emotion. Giving computers more peripherals than a keyboard and a monitor can certainly improve their ability to recognise and express emotions, but such technical advances do not seem to pose a serious theoretical challenge to the internalist view of emotion. The internalist can simply treat all the extra peripherals as providing richer input to, and more expressive output from, the central cognitive system where the 'real' emotional processes go on. All discussions of the ‘recognition’ and ‘expression’ of emotion are perfectly consistent, and perhaps even reinforce, a view of emotions as essentially *brain*-processes. The body just supplies input to the brain or serves as a vehicle for output, but the interruption mechanisms themselves are implemented exclusively in the brain. The input-output boundary is thus firmly maintained between the brain and the rest of the body.

Embodied emotions

Classical cognitive science is not necessarily committed to the internalist view of emotion, and the internalist view of emotion does not logically entail that bodily inputs are non-specific (i.e. that bodily inputs do not determine which emotional response is produced). However, in practice most classical cognitive scientists did, in fact, take an internalist view of emotion, *and* assumed that bodily inputs were non-specific. The most famous proponents of this view were Stanley Schachter and Jerome Singer. On the basis of a well-known experiment on the emotional effects of adrenaline (Schachter and Singer, 1962), Schachter argued that emotions arise from cognitive attributions about one's state of physiological arousal (Schachter, 1964). In other words, all the body does is to tell the brain that *some* kind of emotional response is called for, without specifying *which* kind. The bodily input is merely a quantitative signal which, when it exceeds some threshold, triggers a cognitive process of appraisal such as those described in section 1.2. According to Schachter, it is this cognitive process that determines which emotional process will be set in motion.

This view of the role of the body in emotion goes back to Walter Cannon, who proposed it as an alternative to an earlier view put forward by William James (Cannon, 1927). James had argued, at the end of the nineteenth century, that all conscious feelings associated with emotions are merely the perception of physiological changes (James, 1888). This implies that bodily inputs are specific to each type of emotion. On the basis of some experimental work, Cannon argued that James was wrong, and that bodily inputs did not differentiate the various emotions from each other. Cannon's arguments won the day, and held sway for the rest of the twentieth century, seemingly confirmed by the famous 1962 study by Schachter and Singer. More recent experimental work, however, has led some researchers to call for a return to a more Jamesian view of emotion. Paul Ekman, for example, has published a study purporting to show that there are, in fact, distinctive ANS signatures for some of the so-called 'basic emotions', such as fear and anger (Ekman, Levenson et al., 1983). There are empirical grounds, then, for reviving the debate about the role of bodily inputs in emotional processing. It can no longer be assumed, as it was for much of the twentieth century, that bodily inputs are not specific to certain emotions. If Ekman's data can be replicated, we must broaden our view of where emotional computation takes place; it may not be the brain alone that instantiates this computation, but the brain plus other bits of the body.

The idea here is that the brain may offload some of its computational burden in emotional processes by storing some information about its current or recent emotional states in the body rather than storing it all in the brain itself. In that case, the emotional system consists not just of some neural structure but of the relevant brain structure plus some extra-neural bodily state. Once again, it would be open to the internalist to redescribe this in terms of an emotional brain mechanism monitoring input from the body as well as from the rest of the world. However, but since the interaction between brain and body is two-way, it allows for complex feedback loops that cross the boundaries between brain and body, so the emotional system is better seen as spread out over the whole agent. Andy Clark and Michael Wheeler have dubbed this effect ‘causal spread’.

Thinking along these lines, Dolores Cañamero has written a computer program that simulates some of the physiological aspects of human emotions. The program runs in a two-dimensional virtual world inhabited by virtual creatures called ‘Abbots’ and ‘Enemies’. Enemies do not have emotions, but Abbots have six basic emotions, each of which can be triggered by both external events and by internal physiological changes. The Abbots’ virtual physiology includes synthetic hormones such as endorphins which, as in humans, can trigger a state of happiness or reduce the perception of pain (Canamero, 1997).

Embedded emotions

Perhaps the claim that emotions are embodied is not that surprising. After all, the idea that emotions involve the body as well as the brain is probably more intuitive than similar claims about thoughts. The idea that emotions are embedded, however, is certainly not so intuitive. The claim that emotional processes supervene on bits of the *external world* as well as bits of the organism is so much of a challenge to the traditional internalist view of emotion that it needs to be argued for in more detail.

To illustrate what is meant by the claim that emotions are embedded, I will discuss a number of external devices that humans use to interrupt ongoing activity. Specifically, I will discuss two kinds of such external device: emotional artefacts, and emotional institutions. This is not meant to be an exhaustive taxonomy.

Emotional artefacts.

In the previous section, I mentioned a number of 'cognitive artefacts' such as maps and calculators. People use these devices to ease the burden of cognitive activity by partially offloading it onto the environment. In this section, I want to argue that humans also use a number of 'emotional artefacts'. These are devices that ease the burden of emotional (interruptive) activity by offloading it in a similar way.

(i) Clocks

Clocks may not appear to have much to do with emotion according to our pretheoretic understanding of the term, but if we accept the definition of emotion given in the previous chapter, we may decide we have to revise this initial impression. Clocks clearly function as interruption mechanisms. They may have other functions too, but it can scarcely be doubted that one of the main purposes for which clocks are constructed is to enable us to set time limits for various activities. Alarm clocks are designed to interrupt sleep. Timers are used to interrupt cooking, teaching, and many other activities. True, we may ignore them, but then we may sometimes decide to override our *internal* interruption mechanisms too, such as when a soldier persists in marching towards the enemy despite the feeling of fear that urges him to run away. If we combine the interruption theory of emotion with the leaky mind hypothesis, clocks may be regarded emotional artefacts.

It should hardly be necessary to repeat here that the normal internalist response is always possible. The internalist can simply argue that clocks provide external input into the real emotional system, which supervenes entirely on the brain or some part of it. It should be clear by now how the proponent of the situated approach to emotion can deal with such a response. The behaviour-generating system of interest here comprises the brain, the body and the clock. It is spread out over the agent and his environment. To insist that we must divide this system up along traditional lines by identifying the input-output boundary with the skin of the organism is to give up on the richer explanatory perspective offered by the situated approach. The purely logical possibility of re-analysing the system in internalist terms is not really all that exciting.

(ii) Fire alarms and burglar alarms

Fire alarms and burglar alarms are also clear examples of interruption mechanisms. In particular, they partially offload the computational burden of the fear mechanism onto the outside world. It might be the case that, other things being equal, those who have such alarms installed in their houses tend to show fewer neural or physiological effects of fear in the relevant contexts than those without such alarms. I know of no evidence to support this hypothesis, but if it were true, it would constitute strong empirical support for the situated approach to emotion, for then we would have demonstrated very tight feedback effects between the neural/physiological aspects of the fear system and the external/mechanical aspects. Alternatively, if it could be shown that those with fire or burglar alarms show more neural or physiological correlates of fear in the relevant contexts than those without such alarms, we would also have evidence for the kind of feedback effects that favour a situated approach to emotion. The absence of feedback effects of either kind would not demolish the situated approach to emotion, but it would raise questions about its explanatory power.

(iii) Cruise-control

Some cars are fitted with cruise-control. This device allows the driver to set a maximum speed beyond which the vehicle cannot accelerate. If the driver then tries to drive faster than the pre-set speed limit, the mechanism cuts in and temporarily disables the throttle. This device is clearly an interruption mechanism. Like the fire alarm and the burglar alarm, it eases the computational burden of fear by partially offloading it onto the car. The driver no longer has to worry about going too fast.

Emotional institutions

The emotional artefacts described so far all offload fairly simple emotions like fear. What about other, more social emotions, of the kind discussed by Robert Frank? Can these too be partially offloaded onto the external environment?

As I have already noted in the previous chapter, Frank argues that many social emotions such as love and guilt function as commitment devices. I have already argued that commitment devices are just a special case of interruption mechanism, so I will not repeat the argument now. The important point to note here is that commitment devices can just as easily be offloaded onto the external world as other kinds of interruption mechanism. However, as befits their more social nature, these kinds of emotion are more typically offloaded onto the social environment than onto the physical one. They are offloaded, in other words, not onto physical artefacts, but onto social institutions. Such institutions therefore deserve to be called ‘emotional’.

(i) Marriage

Marriage is the emotional institution *par excellence*. Its function is to offload some of the computation involve in romantic love onto the social environment. According to Frank’s evolutionary analysis, the function of romantic love is to ensure that you can form a stable pair-bond for long enough to raise at least one child. Since this function cannot be fulfilled without a multitude of conditions, romantic love may in fact designate a whole suite of emotions, rather than just one. For example, we might distinguish romantic fidelity from romantic generosity. These are not usually regarded as distinct emotions, but there are many reasons why common sense views of love might be more misleading than many other aspects of folk psychology, so we should not worry too much if the cognitive science of love leads to counter-intuitive findings.

If this view of love is correct, we need not be surprised if we find that marriage is a way of offloading *some* aspects of romantic love but not all. We can take each aspect of romantic love separately, and ask, for each one, whether or not it is plausible to see marriage as a way of offloading it.

As far as romantic fidelity goes, marriage seems to fit the bill as a means for offloading. Once two people have declared their vows in public, they do not always need to interrupt themselves when faced with an offer of extra-marital sex. If there are friends and relatives around, they will probably exert social pressure to make sure the vows are kept.

Marriage is also a way of offloading romantic generosity. The common practice of exchanging gifts at wedding anniversaries means that there is, in effect, an external timing mechanism that repeatedly interrupts the day-to-day routine of married life and prompts the giving of gifts. Here again, the scrutiny of friends and relatives can help to enforce this interruption mechanism.

(ii) Confession

By means of the institution of confession, many Catholics partly offload guilt. The rule forbidding one to take communion while in a state of sin, coupled with the expectation that one will take communion fairly regularly, means that there is in effect a kind of external alarm clock that repeatedly forces one to examine one's conscience. It is easy to see how this institution is a way of offloading the computational load of constantly checking to see if you have kept your promises. When you have not, you must tell a priest, who then exerts social pressure on you to make ammends.

Emotional groups

From what has been said so far, one might easily gain the impression that emotions can only be offloaded onto the environment by creatures like ourselves, who are sophisticated enough to dream up ways of doing so. This, however, is not the case. For those animals with less sophisticated minds, natural selection can do the design work instead. Merekats did not sit down and work out their system of taking turns to do sentry-duty; the system evolved by natural selection, without any conscious reasoning on their part. Yet this too is plausibly described as an embedded emotion. Those merekats that are not doing sentry duty have partially offloaded the computational burden involved in detecting predators onto part of their social environment. They have, that is, partially offloaded the emotion of fear onto another merekat.

Humans, too, offload some emotions onto the social environment without the need for social institutions. The power of crowds to amplify the emotions of each individual member is a case in point. Social psychologists have long remarked on the distinctive nature of emotional behaviour in groups, from the contagious nature of panic to the lynch-mob mentality in which otherwise peaceful people can sometimes be caught up. These too are plausibly described as cases of embedded emotion, when the feedback loops between brain, body and social environment mean that the whole is greater than the mere sum of its parts.

Purely cerebral emotions

Adopting a situated approach to emotion does not imply that all emotions are embodied and embedded. Quite the contrary, in fact. The whole point of the situated approach is not that the input-output boundary is located at some point other than the brain-body boundary; it is that the boundary is moveable. It is not enough to give examples, then, of how emotions can be embodied and embedded; we must also find plausible examples of purely cerebral emotions. Otherwise, the mind would not be *leaky*; it would be permanently *overflowing*.

This is not a problem. There are plenty of examples of emotions where the behaviour-generating system of interest is the bare biological brain. If there were not, the internalist view of emotion would probably not have been so initially tempting. The startle reaction that makes a lone squirrel interrupt its foraging behaviour when it perceives a sudden noise may be an example of such a purely neural emotion. It need not be, however. If the squirrel's extra-cerebral body chemistry has been sufficiently altered by previous startle reactions to warrant being regarded as a repository of emotional memory, then this would be a case of an embodied emotional process. Everything depends on the details of the particular situation. The virtue of the situated approach is that it calls our attention to such details.

Situated emotions and artificial intelligence

If the situated approach to emotion is to count as part of cognitive science, it must lead us to design better artificial emotion systems. I believe it can. In particular it suggests that those working in affective computing do not need to build interruption mechanisms entirely *inside* the machine. When constructing emotional robots, designers can be sensitive to the possibility of exploiting features of the external environment to offload some of the computational burden.

However, the main lesson for affective computing to learn from the situated approach to emotion may be precisely the opposite. If computers seem so unemotional today, this is because they have, in a sense, already offloaded *all* their emotions onto their human environment. Computers do not *need* interruption mechanisms at the moment, because humans take care of all their survival and reproductive requirements. If more interruption mechanisms were installed in desktop computers, so that they shouldered more of the computational burden of caring for themselves, humans would find that their own burden became lighter. Computers would become easier to use.

Things began to move in this direction with the invention of screen-savers, which can thus be seen as the first *internal* emotions in desktop computers. As the example of the screen-saver shows, giving computers emotions can benefit their human users – in this case by allowing monitors to live longer and thus allowing human owners to save money by replacing them less often. Building in other kinds of interruption mechanism may make computers even more user-friendly. For example, a computer that was designed to interrupt activity momentarily when it detected that its user was becoming too tired might help to avoid faulty data being stored in its memory. Computers that become bored when they are not receiving input of a sufficiently novel kind (as in the ACRES program described in section 3.2) might help their users to avoid wasting time. And so on.

These points become even more relevant when the computers in question are designed for entertainment purposes rather than more utilitarian ones.

The simulated agents in gaming software will be more realistic, and therefore more entertaining, when they have more sophisticated emotional capacities. This suggests that developments in affective computing over the next few decades may be driven more by the demands of computer gaming than by the demands of industry or commerce.