Damasio's error

Aaron Sloman

(In The Philosophy Magazine 2004)

In 1994 Antonio Damasio, a well known neuroscientist, published his book *Descartes' Error*. He argued that emotions are needed for intelligence, and accused Descartes and many others of not grasping that. In 1996 Daniel Goleman published *Emotional Intelligence: Why It Can Matter More than IQ*, quoting Damasio with approval, as did Rosalind Picard a year later in her book *Affective Computing*.

Since then there has been a flood of publications and projects echoing Damasio's claim. Many researchers in artificial intelligence have become convinced that emotions are essential for intelligence, and they are now producing many computer models containing a module called 'emotion'.

Before that, serious researchers had begun to argue that the study of emotions and affect had not been given its rightful place in psychology, and cognitive science, but their claims were more moderate. For example, a journal called *Cognition and Emotion* was started in 1987. Even I had a paper in it in the first year.

Damasio's argument rested heavily on two examples. The first was of Phineas Gage. In 1848, an accidental explosion of a charge he had set blew his tamping iron through his head, destroying the left frontal part of his brain:

> He lived, but having previously been a capable and efficient foreman, one with a well-balanced mind, and who was looked on as a shrewd smart business man, he was now fitful, irreverent, and grossly profane, showing little deference for his fellows. He was also impatient and obstinate, yet capricious and vacillating, unable to settle on any of the plans he devised for future action. His friends said he was No longer Gage. (http://www.deakin.edu.au/hbs/GAGEPAGE/Pgstory.htm)

The second example was of one of Damasio's patients,who he refers to as "Elliot". Following a brain tumor and subsequent operation, Elliot suffered damage in the same general brain area as Gage (left frontal lobe). Like Gage, he experienced a great change in personality. Elliot had been a successful family man, and successful in business. After his operation he became impulsive and lacking in self-discipline. He could not decide between options where making the decision was important but both options were equally good. He persevered on unimportant tasks while failing to recognize priorities. He had lost all his business acumen and ended up impoverished, even losing his wife and family. He could

no longer hold a steady job. Yet he did well on standard IQ tests. (See http://serendip.brynmawr.edu/bb/damasio)

Both patients appeared to retain high intelligence as measured by standard tests, but not as measured by their ability to behave sensibly. Both had also lost certain

kinds of emotional reactions. What follows from these cases?

In a nutshell, here is the argument Damasio produced which many people in many academic disciplines enthusiastically accepted as valid:

> Damage to frontal lobes impairs emotional capabilities
>
> Damage to frontal lobes impairs intelligence
>
> Therefore Emotions are required for intelligence

The conclusion does not follow from the premises. (Whether the conclusion is true is a separate matter, which I'll come to.) Compare this argument "proving" that cars need functioning horns in order to start:

> Damaging the battery stops the horn working in a car
>
> Damage to the battery prevents the car starting
>
> Therefore a functioning horn is required for the car to start

A moment's thought should have reminded Damasio's readers that two capabilities could presuppose some common mechanism, so that damaging the mechanism would damage both capabilities, without either capability being required for the other. For instance, even if both premises in the horn argument are true, you can damage the starter motor and leave the horn working, or damage the horn and leave the starter motor working.

I first criticised Damasio's argument in two papers in 1998 and 1999 and have never seen these criticisms of Damasio's arguments made by other authors. My criticisms were repeated in several subsequent publications. Nobody paid any attention to the criticism and even people who had read those papers continued to refer approvingly to Damasio's argument in their papers. Very intelligent people keep falling for the argument. For example, Susan Blackmore did not notice the fallacy when summarising Damasio's theories in her excellent recent book C*onsciousness: An Introduction (2003)*. (She has now informed me that she agrees that the argument used is fallacious.)

The best explanation I can offer for the surprising fact that so many intelligent people are fooled by an obviously invalid argument is sociological: they are part of a culture in which people want the conclusion to be true. There seems to be a wide-spread (though not universal) feeling, even among many scientists and

philosophers, that intelligence, rationality, critical analysis and problem-solving powers are over-valued, and that they have defects that can be overcome by emotional mechanisms. This leads people to like Damasio's conclusion. They *want* it to be true. And this somehow causes them to accept as valid an argument for that conclusion, even though they would notice the flaw in a structurally similar argument for a different conclusion (such as in the car horn example). This is a general phenomenon. Consider, for instance, how many people on both sides of the evolution/creation debate, or both sides of the debate for and against computational theories of mind, tend to accept bad arguments for their side.

A research community with too much wishful thinking does not advance science. Instead of being wishful thinkers, scientists trying to understand the most complex information-processing system on the planet should learn how to think (at least some of the time) as designers of information-processing systems do.

To be fair, Damasio produced additional theoretical explanations of what is going on, so, in principle, even though the quoted argument is invalid, the conclusion might turn out to be true and explained by his theories. However, his theory of emotions as based on 'somatic markers' (regulatory signals in the brain's representation of the body) is very closely related to the theory of William James, which regards emotions as a form of awareness of bodily changes. This sort of theory is incapable of accounting for the huge subset of socially important emotions in humans which involve rich semantic content which would not be expressible within somatic markers (such as admiring someone's courage while being jealous of his wealth) and emotions that endure over a long period of time while bodily states come and go (such as obsessive ambition, infatuation, or long term grief at the death of a loved one).

The key assumption, shared by both Damasio and many others whose theories are different in details, is that all choices depend on emotions, and especially choices where there are conflicting motives. If that were true it would support a conclusion that emotions are needed for at least intelligent conflict resolution.

Although I will not argue the point here, I think it is very obvious from the experience of many people (certainly my experience) that one can learn how to make decisions between conflicting motives in a totally calm, unemotional, even cold way, simply on the basis of having preferences or having learnt principles that one assents to. Many practical skills require learning which option is likely to be better. A lot of social learning provides conflict resolution strategies for more subtle decisions: again without emotions having to be involved. Of course, one could make a terminological decision to label all preferences, policies, and principles 'emotions'. But that would trivialise Damasio's conclusion.

So, let's start again: what are emotions, and how do they work? There are many ways to study emotions and other aspects of human minds. Reading plays, novels or poems will teach much about how people who have emotions, moods, attitudes, desires and so on think and behave, and how others react to them,

because many writers are very shrewd observers. Studying ethology will teach you something about how emotions and other mental phenomena vary among different animals. Studying psychology will add extra detail concerning what can be triggered or measured in laboratories, and what correlates with what. Studying developmental psychology can teach you how the states and processes in infants differ from those in older children and adults. Studying neuroscience will teach you about the physiological brain mechanisms that help to produce and modulate mental states and processes. Studying therapy and counselling can teach you about ways in which things can go wrong and do harm, and some ways of helping people. Studying philosophy with a good teacher may help you discern muddle and confusion in attempts to say what emotions are and how they differ from other mental states and processes.

There's another way that complements these: do some engineering design. Suppose you had to design animals (including humans) or robots capable of living in various kinds of environments, including environments containing other intelligent systems. What sorts of information-processing mechanisms, including control mechanisms, would you need to include in the design, and how could you fit all the various mechanisms together to produce all the required functionality, including: perceiving, learning, acquiring new motives, enjoying some activities and states and disliking others, selecting between conflicting motives, planning, reacting to dangers and opportunities, communicating in various ways, reproducing, and so on?

If we combine this "design standpoint" with the other ways to study mental phenomena, we can learn much about all sorts of mental processes: what they are, how they can vary, what they do, what produces them, whether they are essential or merely by-products of other things, how they can go wrong, and so on. The result could be both deep new insights about what we are, and important practical applications.

The design-based approach is not new: over the last half century, researchers in computational cognitive science, and in artificial intelligence have been pursuing it. Because the work was so difficult, and because of pressures of competition for funding and other aspects of academic life (such as lack of time for study outside one's own specialism), as more people became involved, the research community became more fragmented, with each group investigating only a small subset of the larger whole, and talking only to members of that group.

Deep, narrowly focused, research on very specific problems is a requirement for progress, but if everybody does only that, the results will be bad. People working on natural language without relating it to studies of perception, thinking, reasoning, and acting may miss out on important aspects of how natural languages work. Likewise those who study only a small sub-problem in perception may miss out ways in which the mechanisms they study need to be modified to fit into a larger system. The study of emotions also needs to be related to the total system.

We may be able to come up with clear, useful design-based concepts for describing what is happening in a certain class of complex information processing systems, if we study the architecture, mechanisms and forms of representations used in that type of system, and work out the states and processes that can be generated when the components interact with each other and the environment.

If the system is one that we had previously encountered and for which we already have a rich and useful pre-scientific vocabulary, then the new design-based concepts will not necessarily replace the old ones but may instead refine and extend them. For example, they might lead us to new sub-divisions and bring out deep similarities between previously apparently different cases.

This happened to our concepts of physical stuff (air, water, iron, copper, salt, carbon and soon) as we learnt more about the underlying architecture of matter and the various ways in which the atoms and sub-atomic particles could combine and interact. So we now define water as H_2O and salt as NaCl rather than in terms of how they look, taste or feel, and we know that there are different isotopes of carbon with different numbers of neutrons.

As we increase our understanding of the architecture of mind (what the mechanisms are, how they are combined, how they interact); our concepts of mind (such as “emotion”, “consciousness”, “learning”, “seeing”.) will also be refined and extended. In the meantime, muddle and confusion reign.[1]

Aaron Sloman, School of Computer Science, The University of Birmingham, UK

http://www.cs.bham.ac.uk/~axs

1 I would like to thank Julian Baggini for help in producing this article for The Philosophers' Magazine