TOWARDS AN INFORMATION PROCESSING THEORY OF EMOTIONS

Aaron Sloman
Oct 1992

The “Attention and Affect” project has the following aims:

(1) to identify high level functional requirements for the architecture of intelligent agents like human beings, especially requirements concerned with processing of motives and control of attention,

(2) to explore, at a “coarse-grained”, global level, possible designs capable of fulfilling those requirements,

(3) to implement working models to test and demonstrate the properties of the designs,

(4) to design effective interfaces for demonstrating the key features of the models, and to enable them to be used for teaching ideas potentially relevant to human control systems,

(5) to use the generative power of proposed mechanisms as a basis for constructing a conceptual framework for describing affective states involving control of attention.

Among the phenomena to be explained are emotional states, and related kinds of “affective” states, e.g. desires, moods, attitudes, obsessions, etc.

Before attempting any general characterisation, let’s look at some examples.

Some examples of emotional states

All the following would normally be regarded as examples of emotional states:

(a) Grief at the accidental death of a young child
(b) Being jealous of someone favoured by one's beloved
(c) Embarrassment at being discovered in a ridiculous situation
(d) Feeling guilty about money one has embezzled.
(e) Shame at being exposed as an embezzler
(f) The thrill of being selected for a team

Do these examples have anything in common?

How do they differ from things that are NOT emotional states?, e.g.

(a) Wishing fewer children were killed on the road
(b) Hoping that your colleagues will like you
(c) Wishing that the streets outside your home were cleaner
(d) Wondering when you will be able to pay back a loan
(e) Wanting other people not to know what you've done.
(f) Receiving something you glad of

All the above states involve something like desires or preferences: what I'll loosely call "motivators". All involve cognitive processes in which some thing, person, event or state of affairs is known about, thought about, believed to exist, etc.

I suggest that the **main** feature in first set of states, not found in the second set, is a certain kind of “mental disturbance”. What kind? Conjecture: a key feature is:

A PARTIAL LOSS OF CONTROL OVER ONE’S OWN THOUGHT PROCESSES.

However not all thought processes involving loss of control are emotions: e.g. a bright light or a sound may divert your attention whether you want it to or not.

Incidentally, not all cases of loss of control are undesirable. People sometimes put themselves in situations where they experience uncontrollable thrills, e.g. on a roller-coaster. And much human interaction is aimed at producing pleasurable states of “passion”. So there’s no pejorative intent in the phrase “loss of control”.

WARNING

Debates about what emotions “really” are are stupid and get knowhere. What is important is to find good ways to classify different kinds of mental states and processes, in the light of good explanatory theories. Compare debates about how to define “water” before people knew what underlying mechanisms produced the properties of different kinds of substances.

You can’t have a good theory of what water is without a general theory about a wide range of types of physical substances, how they are produced, how they interact, how their properties change, etc. Similarly you cannot have a good theory about the nature of particular sorts of mental phenomena, e.g. emotions, without having a good theory about a whole range of mental phenomena and the mechanisms that can produce, maintain, modify, or terminate them.

THEORIES ABOUT EMOTIONS

There are many theories of the nature of emotion, varying in the features they regard as central to the concept. E.g.

(1) Some concentrate on physiological processes (sweating, blushing, weeping, muscular changes, etc.), which are then sensed as part of 'feedback' from the body.

(2) Some concentrate on introspective features of different kinds of states: "How does it 'feel' "?

(3) Some (Freud?) concentrate on subconscious mental processes, e.g. subconscious desires, intentions, beliefs, memories.

(4) Some concentrate on allegedly basic sets of emotions and states derived from them, e.g. anger, fear, sadness, joy, etc.

(5) Some claim that emotional states essentially involve consciousness of the state. (As if one could not be upset or angry without being aware of it.) Depending on one's view of consciousness this can lead to anti-scientific theories of emotions and the like.

(6) Some, following H.A. Simon, concentrate on information processing and control mechanisms underlying overt and mental behaviour.

The only way to get a deep understanding of these and other mental phenomena is to view the mind (or if you prefer, the brain) as a very sophisticated control system, and try to discover what control tasks it has to perform, what constraints there are on the performance of those tasks, and then what kinds of mechanisms are capable of performing those tasks.

This is a multi-disciplinary investigation including theoretical analysis of the nature of various kinds of control requirements, a study of possible designs, empirical research concerning what kinds of mental processes actually occur, and neurobiological research into the underlying machinery.

All of these are extremely difficult investigations. In all areas our current state of knowledge is abysmal, and researchers of the future looking at our research publications will treat them much as we treat the work of alchemists trying to understand the behaviour of physical matter. But their early gropings laid some of the foundations for important later work. So can ours, I hope.

The theoretical analysis of design requirements and the forms of mechanisms that can meet those requirements, is much like engineering, and in some ways like traditional philosophy. It needs to be informed by empirical studies of the phenomena in question and the mechanisms supported by the brain. But empirical studies not informed by a deep theory will usually tell you very little. (They tell you most when you have two deep rival theories with conflicting predictions.)

“Bottom Up” vs “Top Down” Research

It’s worth noting that the analysis of requirements, designs and mechanisms can be bottom up or top down: i.e. you can either

(a) study fragmentary mechanisms and try to find out what happens when they are combined and see whether any combinations can produce the sorts of capabilities you wish to explain,

OR

(b) study global requirements and try to form a global architecture, then work down towards possible underlying mechanisms.

Usually neither approach works on its own. So a mixture of bottom up, top down, and middle out collaborative research is required. My own emphasis is mostly top down: but that’s why it is important for me to collaborate with others who do it differently.

WHY “INFORMATION PROCESSING?”

There are many different kinds of control systems, e.g. homeostatic systems in the body and many designed by engineers. Most of them are (a) quantitative (b) direct, in that the things being controlled are directly causally connected with the controlling mechanisms and produce direct feedback.

By contrast, in intelligent agents, much of the information that is needed for deciding what to do, how to do things, how to resolve conflicts of preferences, etc. involves information that is not quantitative, i.e. best

represented by sets of numbers, but is far more "structural", i.e. best represented by things like sentences, diagrams, maps, networks, and other forms of symbolism. This is not quantitative. (Of course, some aspects of how we behave are quantitative: but they are only a subset.)

Also much of what intelligent agents are trying to achieve, avoid, maintain, is not concerned with physical states that they can directly and continuously control. Rather it is often concerned with the distant future, absent objects, events that might occur but haven't yet. This means that the control has to go via "representations" of those things, i.e. information structures with semantic relationships to other things.

Investigation of control systems with these two features (partly) non-quantitative, and (mostly?) semantic require concepts and techniques that so far have been developed within the study of information processing rather than the study of electronics, chemistry, physics, (traditional) psychology, physiology, etc.

“Embarassment”, “Shame” and “Guilt”

What do these have in common? They all involve a wish that something had not happened.

Embarrassment at being discovered doing something

This (normally) involves:
(a) believing other people are aware of one’s situation
(b) believing they have certain thoughts and judgements about the situation
(c) wishing that they were not aware of it
(d) wishing that they did not have those thoughts

Note that it need not involve wishing one had not done it.

Feeling guilty about money one has embezzled.

For this it is not necessary that others have any information about what has happened. It does involve:
(a) wishing one had not done, whatever it was
(b) the wishing is not based only on fear of discovery, or concern about bad consequences for oneself, but consideration of some “higher” or “external”, or “objective” standard, which would be equally applicable to other people.

These states can include conflicts with other motivators, e.g. pleasure or joy, or relief at what one was able to achieve as a result of the embezzlement, e.g. paying off the blackmailer, paying for the operation that cured one’s child’s paralysis: many emotional states involve a mixture of emotions of different sorts.

Shame at being exposed as an embezzler

This seems to involve combinations of the previous two: i.e. genuine wishing that it had not been done, and not only for selfish reasons, plus a concern about the knowledge and opinions of others.

However you can feel shame without guilt, when others wrongly believe you have done something that if you had done it would have made you feel guilt.

One can feel shame even though others have not discovered what happened: then the state is very close to guilt. It's something like shame before oneself.

NOTE: these are not intended to be complete analyses. They are merely indications of the cognitive and motivational complexity involved in certain kinds of states that we often talk about, without analysing what we mean.

The points made so far fail to account for what it is about these states that makes them "emotional". It is possible to satisfy all the descriptions so far without being in any way upset, disturbed, moved, etc. I.e. without being emotional.

Emotionality commonly involves something else: that's where partial loss of control comes in. It can be extreme, as in hysteria or obsession, or slight, we need a theory of the mechanisms that produce such states.

TOWARDS A THEORY...

Multiple sources of motivation

Hierarchies of control

Concurrency

Resource limits

The need to prevent excessive diversion of attention

The need to allow diversion in special cases

The need for the decisions to be fast and simple

The impossibility of combining all these in something that always works perfectlly.