The Terminological Pitfalls of Studying Emotion

(In the process of being updated)

Tim Read* and Aaron Sloman
Cognitive Science Research Centre
School of Computer Science
University of Birmingham
Birmingham, UK
Email: T.M.Read@cs.bham.ac.uk, A.Sloman@cs.bham.ac.uk

August 28, 1993

Abstract

The research community is full of papers with titles that include terms like ‘emotion’, ‘motivation’, ‘cognition’, and ‘attention’. However when these terms are used they are either considered to be so obvious as not to warrant a definition, or are defined in overly simplistic and arbitrary ways. The reasons behind our usage of existing terminology is easy to see, but the problems inherent with it are not. The use of such terminology gives rise to a whole set of problems, chief among them are confusion and pointless semantic disagreement.

These problems occur because the current terminology is too vague, and burdened with acquired meaning. We need to replace it with terminology that emerges from a putatively complete theory of the conceptual space of mechanisms and behaviours, spanning several functional levels (e.g.: neural, behavioural and computational (Gray, 1990)). Research that attempts to use the current terminology to build larger and more complex theory, just adds to the existing confusion.

In this paper I examine the reasons behind the use of current terminology, explore the problems inherent with it, and offer a way to resolve these problems. The days when one small research team could hope to produce a theory to explain the complete range of phenomena currently referred to as being ‘emotional’ have passed. It is time for concerted and coordinated activity to understand the relation of mechanisms to behaviour. This will give rise to clear and unambiguous terminology that is defined at different functional levels. Until the current terminological problems are solved, our rate of progress will be slow.

*This paper is written by the first author with ideas developed from conversations with the second. This research is supported by the SERC, studentship number: 91313564, and by the MRC, grant number: SPG9200393

1 Introduction

If the reader of this paper were to pick up a handful of papers that were on the subject of ‘emotion’, they would probably find them full of references to the word ‘emotion’, or supposed instances like ‘happy’, ‘sad’, ‘anxious’, ‘curious’, ‘contented’ etc. However when these terms are used they are often not defined (the author assuming that the definitions are either so obvious or so well known that they are not worthy of explanation), or if they are defined, the definitions produced are overly simplistic and arbitrary.

The study of mental architectures has reached the stage where there needs to be coordinated research activity on a large scale to make real progress in the sorts of problems we are trying to solve. Such collaboration will need to bring together people from different backgrounds, such as neuroscience, psychology, biology, computer science etc. However before such enterprises can succeed, there are still terminological problems that need to be resolved. The problems I am referring to are the ones associated with generating and using definitions before the research is ready to support them. This will lead to confusion, unnecessary argument, and a duplication of effort and resource.

The terminological problems are centred around the use of such terms as ‘attention’, ‘cognition’, ‘motivation’, ‘emotion’, and ‘consciousness’. This paper is split into four sections following this one that detail these problems. The first section attempts to understand why we have our current terminology, the second looks at the problems with using it, and the third offers a way to resolve the problems. Finally there is a conclusion that draws out the salient points from the paper.

2 Why Do We Use the The Current Terminology?

Our current terminology is well used within the research community, both the terms themselves, and the way in which they are used. A little consideration of the area will illustrate why this use of terminology occurs:

- It enables researchers to place their work in the field. E.g.: X is a paper on ‘emotion’, “I am working on the brain systems that underlie ‘emotion’”.
- It is used in an attempt to allow researchers from different disciplines to reach common ground (a mistaken belief, as will be shown later).
- It is much easier to use a simple term than do the hard job of trying to describe exactly what it corresponds to in terms of mechanisms and behaviour.
- Researchers grew up on the language as undergraduates etc, and hence carry it forward into the research arena where it is both expected and encouraged.
- It can be used as a carpet, under which the ambiguities and unknowns of the work can be swept. This places the burden of understanding a piece of work on the shoulders of the reader. This is not a satisfactory state of affairs.

Another factor is that even if a researcher is aware of the problems with the terminology they are using, it can be very difficult to produce new terminology. It may appear to a researcher to be a diversionary task, from the *real* work of producing a theory about the subject at hand. The terminology is seen simply as a tool for communicating the theory. The fallacy with this approach is looked at in the next section. It is interesting to note that given the large quantity of literature claiming to offer theories of 'emotion', there is still no clear consensus as to what exactly is 'emotion'.

3 What Are The Problems With The Current Terminology?

It has been said that there are as many theories of 'emotion' as there are theorists (Scherer, 1993). The same terms are often given different definitions by researchers. To some extent definitions are arbitrary and not subject to being proven (Scherer, 1993). However it is important to note the distinction between the conceptual space of mechanisms and behaviours be explored, and how we refer to them (Sloman, 1993a). The former is what we as researchers are interested in explaining, the later is the terminology with which we choose to do it. If we choose arbitrary terminology which is overly general, or isn't grounded in the former space, it will lead to problems. Areas of research that either do not take account of this distinction, or work purely at at terminological level will consequently not succeed at producing the new science that will leave outdated muddled concepts like 'emotion' behind.

An example of such a purely terminological approach to understanding and classifying 'emotion' is illustrated by Wierzbicka (1992). The work is an attempt at understanding 'emotion' or 'emotional concepts' by forming semantic primitives that can be used to classify other language that used about supposed particular instances of 'emotion'. An example of Wierzbicka (1992)'s work is:

> The main difference between *happy* and *joy* consists in the personal character of the former and the unspecified, possibly impersonal, character of the latter...
>
> Furthermore, unlike *joy*, and unlike *unhappy*, *happy* implies an absence of further desires...

As an attempt to study language, this work may have something useful to say. However it doesn't move us any nearer to understanding what is meant by 'emotion'. The above use of the terms *happy* and *joy* are open to different interpretation than that put forward by Wierzbicka (1992). The way to resolve these disagreements are to have a complete theory of the mechanisms and behaviours, from which appropriate terminology will arise naturally as explained below.

Even when the conceptual space/terminology distinction is understood, there is still room for terminological error to arise. Sometimes a definition will be produced which only fits a small set of the phenomena associated with the concept by the research community as a whole. The phenomena which don't fit the definition are either ignored or dismissed as being epiphenomena, or as having little or no importance. Problems arise when researchers try to collaborate and discuss their work. Instead of being able to focus on the content and consider the similarities and differences, the discussion can reduce into a disagreement over terminology, what it means and who is right.

Often a definition of 'emotion' will be given in an extremely simplistic terms, in a few words or a paragraph at most. These definitions can often be split into two categories:

- Conjunctive: A + B + C + .. Such definitions can be falsified by considering an example that doesn't contain all the set elements. An example of such a definition is the James-Lange theory of emotion (autonomic arousal + labelling). (Atkinson et al., 1983: Summary).
- Disjunctive: A or B or C or .. Such definitions can be so broad as to tend to blur the boundaries between example and non example phenomena. An example of such a definition can be illustrated by considering the argument that could arise over whether terms like 'anger' and 'curiosity' are examples of 'emotional phenomena'.

In both of the above categories the definitions produced are too vague to reflect accurately all the facets of phenomena often referred to as being 'emotional'. An example of the scope of such phenomena can be seen by looking at the book by Frijda (1986).

Another problem that emerges because of our current terminology is the idea of a 'basic set of emotions' (Oatley and Johnson-Laird, 1987; Lazarus, Kanner and Folkman, 1980; Fisher, Shaver and Carnochan, 1990: amongst others). There is much argument against this position (e.g.: (Ortony and Turner, 1990)). In the context of the arguments put forward in this paper, the notion of a 'basic set of emotions' is fundamentally untenable. I.e.: If you don't have a complete theory of the mechanisms and behaviours that correspond to phenomena currently crudely referred to as being 'emotional', you aren't in a position to classify supposed instances of it into sets. A similar argument holds against the simplistic notions that some 'emotions' are the opposite of others (often shown as the labels for axes on graphs), and that 'emotions' can be graded by simple criterion (e.g.: +ve and -ve (Lazarus, Kanner and Folkman, 1980)).

A related terminological problem is evident in the argument between Zajonc and Lazarus on the 'primacy of cognition in affect' (LeDoux, 1987; Scherer, 1993). As LeDoux (1987) points out:

> The debate between Lazarus and Zajonc largely amounts to a question of how certain aspects of information processing by the brain should be viewed.

Or to put in the context of the arguments in this paper, their argument is itself a problem due to the terminology they are using. As neither of the terms 'cognition' or 'affect' have precise definitions (derived from the mechanisms and behaviour to which they refer), it is not possible to argue about the 'primacy of cognition in affect'. Both researchers produce data to support their respective positions, so attempting to solve the problem at this level just leads to deadlock and confusion.

A similar problem is also highlighted by (LeDoux, 1987) concerning the incorporation of animal data into work on 'emotion'. This is a very contentious subject, the varying positions being put forward depending on the background and dogma of any particular researcher. These positions take their support from the terminology that underpins them. If the terminology is rejected, and the focus returned on the conceptual space of mechanisms and behaviours, the division between animals and humans disappears. Instead there becomes a varying range of behavioural phenomena, supported by architectures of varying complexity and functionality.

A related point to the previous one concerns the problems of anthropomorphising about the way in which proposed architectures operate (McDermott, 1981). This criticism is particularly aimed at researchers from a computer science and Artificial Intelligence background, but also holds for psychologists and neuroscientists who propose little models with titles including terms like 'emotion' and 'motivation' etc. The point to bring out here (that go beyond McDermott (1981)'s criticisms), is that anthropomorphising is itself a problem inherent in the way in which current terminology doesn't have definitions that are grounded in a parsimonious mechanistic theory. Instead of meaningless terms like 'emotion', there needs to be a whole hierarchy of terms, clearly grounded and defined from a mechanistic standpoint. When we reach this point, any such anthropomorphising will be immediately obvious, and consequently be avoided.

4 How Can We Overcome the Terminological Pitfalls?

The pitfalls of using our current terminology have been outlined in the previous section. In order to avoid these problems in the future we need to take care in our use of such terms as 'emotion', 'cognition', and 'attention' etc. In their place we need new terminology that is not burdened with the acquired meaning of many generations, but is grounded in good explanatory theory. Such theory needs to be based on a detailed understanding of the functionality of the underlying mechanisms. Sloman (1993a) states this as follows:

> The notion that science starts with definitions is quite wrong. The definitions can only come after you have good explanatory theories: for only in terms of the theories will you be able clearly and unambiguously specify boundaries between different sets of phenomena.

An example of the role of good explanatory theory in the development of terminology can be seen in the development of atomic theory and the formation of the periodic table of the elements. It was previously thought that the physical universe was composed of just four basic substances: water, fire, air and earth. All the materials that existed in the world could hence only be classified in these simple terms, or by virtue of certain gross properties: texture, colour, hardness, brittleness, weight per unit size, solubility, combustibility etc. It was not until the development of the atomic theory, that matter was finally classified into the basic elements, which make up the periodic table. This groups the basic substances (elements) that consist of atoms of only one kind, in terms of their subatomic components. All matter is composed of these elements, and a particular material's properties can be predicted by considering the elements of which it is composed, and how they interact.

As the atomic theory was being developed, and the periodic table was being formed, new terminology was appearing that was grounded in the theory. This new terminology (atom, proton, neutron, electron, atomic weight, atomic number, element, compound etc), very clearly and unambiguously defined the composition of matter.

Another example is the terminology associated with water (Sloman, 1993a):

> Before 'water' was 'scientifically' defined, it would have been associated more or less loosely with visual appearance, behaviour when poured, taste, effects of

> drinking it, and other characteristics. In such a pre-scientific state, people will waste their time arguing over which of the associations are the 'essential' defining characteristics and which not. E.g.: If something looked, tasted, and flowed like water, but killed those who drank it some might say it wasn't water, because being good to drink is part of their *definition* of water, while others might argue that it was poisonous water because it possessed all the features *they* thought were part of the definition of 'water'.

We are at a similar position with our current use of terminology like 'emotion' and 'cognition'. Trying to use such language as building blocks from which to form theory, and other definitions will just lead to confusion and pointless semantic disagreement. It is still useful to use terms like 'emotion' as keywords (in a non scientific sense), to enable easy identification and placement of research work, but not as a scientific concept (cf Sloman (1991)'s arguments on consciousness).

Definitions need to be produced based upon putatively complete theories of the subject in hand, based on a detailed understanding of the functionality of the underlying mechanisms. Such theory needs to explain the operation of the mechanisms at different levels, such as the three levels Gray (1990) argues for his model: Behavioural, neural, and computational. It may turn out to be the case that more than three levels are eventually required to support a complete theory (cf the layers of a virtual machine).

Furthermore as we are currently a long way from having a sufficient understanding of the mechanisms to be able to produce theories rich enough to generate good definitions, we need to be even more restrictive in our use of terminology. We need to list the specific phenomena we are studying and keep our theories to the mechanisms we believe are responsible for them.

Producing terminology from a theory that describes the mechanistic functionality and maps across all of Gray (1990)'s three levels will not be an easy task. The reason being that the mappings of different functional concepts will not necessarily be equally complex across the three levels. Hence working toward a complete theory which will produce such mappings, will produce terminology en-route which may initially exist at only one or two levels, and have no equivalent mapping at the third. This is a necessary step in the process of producing a mapping of functionality across all three levels, and will be resolved when the theoretical concepts have been developed for the other levels. An example of such terminology from our own group here there are terms like 'insistence' and 'urgency' (Sloman, 1993b). These definitions are grounded in a theory about a virtual machine architecture for resource limited agents (Sloman, 1993b), (Sloman, 1987), but have meanings only at Gray (1990)'s computational and behavioural levels. For them to become part of a complete theory, they will need to be mapped onto a neural level (in which they may correspond to a whole series of circuits with associated terminology).

Once we start to produce terminology that comes from a theory that refers to specific behaviours and mechanisms (defined at different functional levels), it will be possible to compare different terminology in terms of the mechanisms and phenomena to which they refer. This will allow us to falsify or extend them as appropriate. Gradually larger and more complex theory can be built that span Gray (1990)'s three levels, which eventually lead to terminology that replace such things as 'cognition' and 'emotion' in an exact and parsimonious way. When this happens the current arguments about existing terminology will resemble the arguments that predated atomic theory, as illustrated by the establishment of the periodic table of the elements, or the understanding of the properties of water.

5 Conclusion

This paper has highlighted the problems with current terminology such as 'emotion'. It offers a way to resolve those problems which will go beyond resolving simple terminological arguments, and work toward a theory of the mechanisms and behaviours evident in life around us.

Up to now the emphasis in research has been for small research teams to spend long periods of time working toward their own particular 'theory of emotion'. This search for the Holy Grail of theories, that will simply and parsimoniously explain what is meant by 'emotion' is misguided. A vague concept like 'emotion' does not refer to a simple set of phenomena that can be explained away by a little concerted effort. What is needed is concerted effort by research teams in a coordinated manner, to provide a theory detailing the functional analysis of mechanisms that are defined at different functional levels.

The terminology that emerges from these research teams, en-route to producing such a theory will be easy to discuss and refine, as it will be based in the conceptual space of mechanisms and behaviour, and will not be bound up in the confusion and rhetoric of current nebulous terminology.

References

Atkinson, R. L., Atkinson, R. C., and Hilgard, E. R. (1983). *Introduction To Psychology, Eigth Edition*, chapter 11, pages 337–339. Harcourt Brace Jovanovich Inc.

Fisher, K. W., Shaver, P. R., and Carnochan, P. (1990). How emotions develop and how they organise development. *Cognition and Emotion*, 4(2):81–127.

Frijda, N. H. (1986). *The Emotions*. Cambridge University Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion*, 4(3):269–288.

Lazarus, R. S., Kanner, A. D., and Folkman, S. (1980). Emotions: a cognitive-phenomenological analysis. In Kellerman, R. P. . H. (Ed.), *Emotion: theory, research and experience. Vol 1 Theories of Emotion*, pages 189–217. San Diego, CA: Academic Press.

LeDoux, J. (1987). Emotion. In *Handbook of physiology, a critical, comprehensive presentation of physiological knowledge and concepts, Section 1, The nervous system*, chapter 10, pages 419–459. Bethesda (Md.), American Physiological Society.

McDermott, D. (1981). Artificial intelligence meets natural stupidity. In Haugeland, J. (Ed.), *Mind Design, Philosophy, Psychology, Artificial Intelligence*. MITP.

Oatley, K. and Johnson-Laird, P. (1987). Toward a cognitive theory of emotions. *Cognition and Emotion*, 1(1):29–50.

Ortony, A. and Turner, T. (1990). What's basic about basic emotions ? *Psychological Review*.

Scherer, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion*, 4(3):1–41.

Sloman, A. (1987). Motives, mechanisms, and emotions. *Cognition and Emotion*, 1(3):217–234.

Sloman, A. (1991). Why consciousness is not worth explaining. In *Proceedings International Conference on Cognitive Science (ICCS91) San Sabastian*.

Sloman, A. (1993a). Personal communication.

Sloman, A. (1993b). Prolegomena to a theory of communication and affect. In A. Ortony, J. Slack, O. S. (Ed.), *To appear in: Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, pages 229–260. Heidelberg, Germany: Springer. Currently available as a Cognitive Science Research Paper, No 194, University of Sussex, UK.

Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science*, 16:539–581.