

# Memory and Emotion in the Cognitive Architecture

W.F. Clocksin

June 26, 2003

Note: This document is an outline that provides some extracts and main points to give some idea of the chapter.

## 1 Introduction

This paper explores issues in memory and affect in connection with possible architectures for artificial cognition. The work described in this paper represents a departure from the traditional ways in which memory and emotion have been considered in AI research, and is informed by two strands of thought emerging from social and developmental psychology. First, there has been an increasing concern with personhood: with persons, agency and action, rather than causes, behaviour and objects [14]. Second, there is an emphasis on the self as a social construct [4], that persons are the result of interactions with significant others, and that the nature of these interactions is in turn shaped by the settings in which these interactions occur [7].

*More here.*

## 2 Emotion

A new field is forming in computer science, named affective computing, defined as “computing that relates to, arises from, or deliberately influences emotions” [13]. The premises of affective computing have not been universally accepted within the AI research community. To understand why this is, we may identify three stages of the understanding of emotions in the context of AI research. These stages are more than chronological trends. More importantly, they are mindsets that describe and constrain the ways that researchers consider to be the role of emotion in intelligent systems.

The first stage, represented by ‘Old Fashioned AI’, holds the traditional dualist distinction between reasoning and the emotions. Because intelligence is seen as based on logical reasoning or abstract problem solving, emotions are undesirable and can only lead to distraction or error. This stage is also expressed by popular science fiction, in which supreme reasoning beings are

free of emotions: The android<sup>1</sup> Mr Data (in the television series *Star Trek: The Next Generation*) has been designed without emotions, and Mr Spock (of the original *Star Trek* television series) is trained from birth to suppress emotions in the service of the logical thinking prized by his culture.

The second stage considers emotion either as a tool or as a generated side-effect of cognitive processes. This is a functionalist approach, in which emotions are seen as useful capabilities for making machines easier to use [13], or for making cognition more efficient – for example – by enabling the intelligent system to react more quickly to threats [15]. The second stage is also expressed in *Star Trek: The Next Generation*, where Mr Data has an emotions chip he can insert into his positronic brain when it is necessary to experience emotions to better perform some task. Most of the time he prefers not to use the chip, regarding emotions as useful in some situations but not necessary.

In the third stage, emotions are seen as the foundation on which all cognition is built. It starts from the premise that intelligence manifests itself only relative to specific social and cultural contexts [17], and is informed by recent progress in social psychology. Emotions constitute an ever-present substrate or foundation to everyday intelligent behaviours. Because cognition and emotion are seen to be constructed within a social context [6, 1], they are intrinsically linked. Seeing emotions in terms of a functional purpose may be useful anthropology, but is not the point: instead, intelligence and meaning is constructed within social relationships, and emotions are the uniquely human capability that enables complex social relationships to happen and for relationships to be expressed through conversation and narrative. The third stage is also expressed in popular science fiction by the films *The Bicentennial Man*, in which a robot develops over a 200-year timespan to become fully human as a result of his relationships with successive generations of a family; and *AI: Artificial Intelligence*, in which a ‘true’ intelligent android is distinguished from other androids by his capacity to love and be loved.

*More here.*

### 3 Memory

From an experiential perspective, memory is often thought of as episodes in which strong feelings of recall briefly dominate a person’s awareness. From a functional perspective, memory is seen as a capability for storage and retrieval of data. However, we consider memory in a broad sense that pervades experience, communication and cognition. Furthermore, it is difficult to justify separating memory from emotion. We have feelings about what

---

<sup>1</sup>An android is an artificial humanoid robot that is nearly indistinguishable from a human in appearance and behaviour.

we remember. These feelings strongly influence the way we plan and learn. Feelings affect what we remember, and how we remember what it is that we remember.

Memory consists of much more than what can be understood from simplistic local measurements of recall. The experience of memory forms parts of ongoing interactions that are emotionally charged and are embedded in a broad social context. Significant memories are to a large extent social phenomena that take place in specific cultural settings. A memory is never a single state: it is an experience.

Human memory relates to so many complex interaction that it is possible that individual memory ‘traces’ may never be detected. Another way of considering the problem of memory is to imagine scenarios in which memory constructs an interactive involvement rather than simply a storage and recall pattern. Support for this way of thinking about memory comes from three different directions: (a) at the psychological level, the social construction of memory, in which memory is seen as a social practice [12, 10], (b) at the physiological level, memory as a continual process of reconsolidation [11], and (c) at the neural network level, where new models based on chaotic dynamics represent memory as nearly periodic orbits that can be altered by small fluctuations [2, 3]. These three independent approaches seem to converge to a common model for memory that is quite different to traditionally held models of memory. The following subsections outline these three convergent approaches.

### 3.1 Social Construction

There are two general strands of research about social construction of memory. Both strands have in common the premise that human memory can only function within a collective context. The first strand is traced back to the work [5] of Maurice Halbwachs (1877-1945), and is primarily concerned with collective memory over a historical timescale: how to we use our experience of the present to reconstruct our historical past, and how this reconstructed past is used to guide our behaviour in the present. Collective memory is selective: different groups of people have different collective memories, which in turn give rise to different modes of behavior. Halbwachs shows, for example, how wealthy old families in France have a memory of the past that diverges sharply from that of the *nouveaux riches*, and how working class constructions of reality differ from those of their middle-class counterparts. The second strand of research is traced back to Lev Vygotsky (1896-1934) who [16] with Alexander Luria emphasized the social dimensions of knowledge construction: relations between people are fundamental to all higher mental functions. This strand tends to focus on the timescale of the individual life: how our memories are formed and changed through our relationships with others. The more recently explored phenomena of

‘false memory’ (e.g. [8]) and the ‘misinformation effect’ [9] can be addressed from a social construction perspective.

*More here.*

### 3.2 Reconsolidation

This subsection reviews Nader [11].

### 3.3 Chaotic Neural Models

This subsection reviews APO memory models [2, 3].

## 4 Conclusion

Because memory and emotion are evolutionarily and developmentally rooted in social interdependence, a thorough exploration of these issues is necessary for the principled design of the cognitive architecture of true intelligent systems. In particular, we have focused on two aspects of the cognitive architecture: emotion and memory. Seeing emotion not as an optional extra, nor as discrete states or episodes of feelings, but as the underlying substrate enabling the formation of social relationships essential for the construction of cognition. Seeing memory not as the storage and retrieval of immutable data, but as a continuous process contingent on experience and never fully fixed or immutable.

*More here.*

## References

- [1] W.F. Clocksin. Artificial intelligence and the future. *Philosophical Transactions of the Royal Society A*, 361:1–28, 2003.
- [2] N.T. Crook and T. olde Scheper. Adaptation based on memory dynamics in a chaotic neural network. *Cybernetics and Systems*, 33(4):341–378, 2002.
- [3] N.T. Crook, T. olde Scheper, and V. Pathirana. Self-organised dynamic recognition states for chaotic neural networks. *Information Sciences*, 150:59–75, 2003.
- [4] K. Gergen. *An Invitation to Social Construction*. London: Sage, 1999.
- [5] M. Halbwachs. *On Collective memory*. University of Chicago Press, ET 1992.
- [6] R. Harre and W.G. Parrott. *The Emotions: Social, Cultural and Physical Dimensions*. Sage, 1996.

- [7] G. Levine. *Constructions of the Self*. Rutgers University Press, 1992.
- [8] E.F. Loftus. Creating false memories. *Scientific American*, 227(3):70–75, September 1997.
- [9] E.F. Loftus and H.G. Hoffman. Misinformation and memory, the creation of new memories. *Journal of Experimental Psychology: General*, 118(1):100–104, 1989.
- [10] D. Middleton and D. Edwards. *Collective Remembering*. Sage, 1990.
- [11] K. Nader. Memory traces unbound. *Trends in Neurosciences*, 26(2):65–72, 2003.
- [12] U. Neisser and R. Fivush. *The Remembering Self: Construction and Accuracy in the Self-Narrative*. Cambridge University Press, 1994.
- [13] R. Picard. *Affective Computing*. MIT Press, 1997.
- [14] J. Shotter and K.J. Gergen. *Texts of Identity*. London: Sage, 1989.
- [15] A. Sloman. Beyond shallow models of emotion. *Cognitive Processing*, 2(1):177–198, 2001.
- [16] L.S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, ET 1978.
- [17] J. Weizenbaum. *Computer Power and Human Reason*. MIT Press, 1976.