Chapter 7.3
Memory and Emotion in the Cognitive Architecture

William F. Clocksin
Oxford Brookes University, UK

ABSTRACT

This chapter explores issues in memory and affect in connection with possible architectures for artificial cognition. Because memory and emotion are evolutionarily and developmentally rooted in social interdependence, a new understanding of these issues is necessary for the principled design of true intelligent systems. We treat emotion not as an optional extra or as a brief episode of feeling, but as the underlying substrate enabling the formation of social relationships essential for the construction of cognition. We treat memory not as the storage and retrieval of immutable data, but as a continuous process contingent on experience and never fully fixed or immutable. Three converging areas of research are identified that hold some promise for further research: social constructionism, reconsolidation memory theory, and memory models based on the nonlinear dynamics of unstable periodic orbits. We argue that the combination of these ideas offers a potentially more substantive approach to understanding the cognitive architecture.

INTRODUCTION

This chapter is concerned with memory and affect in connection with possible architectures for artificial cognition. The work described in this chapter represents a departure from the traditional ways in which memory and emotion have been considered in artificial intelligence (AI) research and is informed primarily by two strands of thought emerging from social and developmental psychology. First, there has been increasing concern with personhood: with persons, agency, and action, rather than causes, behaviors, and objects (Shotter & Gergen, 1989). Second, there is an emphasis on the self as a social construct (Gergen, 1999), 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 (Levine, 1992).

There is a tradition in cognitive science of teasing the cognitive architecture apart into components, so that their features and characteristics may be better understood. Examples of such components are memory, perception, problem solving, attention, and emotions. The reductionist method of decomposing a complicated problem into subproblems that can be studied individually has a long and distinguished history and has well served the physical sciences. However, it is important to remember that cognition takes place within communities and cultures and is found within a complex matrix of interactions and reciprocalities of needs and desires. Although the nominalist philosophy that has been the prevailing thoughtway for more than 300 years has privileged the individual mind as the centre of being and the locus of meaning-producing processes (Churchland, 1995), it is well to remember Weizenbaum's premise (1976) that intelligence is manifested only within a cultural and social milieu. As Aubé concludes elsewhere in this book, it is incoherent to try to design emotional systems for artifacts that do not belong to communities.

It is possible that an understanding of cognition rich enough to provide a computational basis for artificial intelligence will need to depend on understanding the human person, which in turn, depends on understanding the propinquity of interaction between persons. The reductionist and nominalist projects do not offer adequate means of obtaining a conceptual grasp of the complex ecosystem of interactions within communities of persons. As Fraser Watts (2000) said, the human emotions illustrate very well the way in which the biological and social aspect of personhood are intertwined. Emotions have both biological and social aspects, and any attempt to explain one aspect without the other leads to an impoverished account.

The main objective of this chapter is to show how ideas from three different research areas might be caused to converge in order to form a new way to think about cognition. The motivation for forcing this convergence is the need to look beyond the Jamesian tradition of understanding emotion and memory as relatively brief episodes of an individual's experience. Instead, memory and emotion are seen here as partners that coconstruct an interactive involvement between persons. This involvement is extended in time and has a character that is provisional, circular, changing, and never finished. This involvement can be pictured as a rope made of three different strands of research. The first strand is social constructionism (Gergen, 1999). Although most work in the area has taken place only within the past 20 years, its emphasis on the self and meaning as achievements emerging from social interactions can be traced back to G. H. Mead (1934). Social constructionism can provide a way to break out of the habit of thinking of cognition as a solely individual affair governed by the physical processes within an individual's brain, and it emphasizes the circular relationship between interpretation and understanding. It is important to understand this circular relationship as being supported and enabled by brain functions, and this motivates the inclusion of the second and third strands. The second strand is neurobiology, from which a new memory model called reconsolidation has been proposed (Nader, 2003). Reconsolidation is the hypothesis that a memory can be returned to a mutable, sensitive state in which it can be modified, strengthened, changed, or even erased. New studies support a reconsolidation theory, in which the traditional distinction between long-term memory and short-term memory must be reconsidered. The third strand is a branch of nonlinear dynamics referred to as chaotic computation, from which has recently emerged a mathematical model describing how memories may be represented as unstable orbits within a phase space, thus providing theoretical underpinnings for a mutable memory model (Crook & olde Scheper, 2002). In this approach, memories are not stored as distributed patterns of weights in a network of neural units, as in well-known connectionist models. Instead, memory states emerge from the dynamics of the network. The appeal of this model is that the self-organized representations that emerge are a consequence of the resolution of internal and external dynamic