Emotional Circuits and Computational Neuroscience

Jean-Marc Fellous, Jorge L. Armony and Joseph E. LeDoux

Jean-Marc Fellous
Computational Neurobiology Laboratory
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road
La Jolla, CA 92037
phone: (858) 453-4100, x 1618 - fax: (858) 587-0417
email: fellous@salk.edu

Jorge L. Armony
Institute of Cognitive Neuroscience
University College London
17 Queen Square,
London WC1N 3AR, UK
email: j.armony@ucl.ac.uk

Joseph E. LeDoux
Center for Neural Science
New York University
4 Washington Place
New York, NY 10003
email: ledoux@cns.nyu.edu

Introduction

Emotion is clearly an important aspect of the mind; yet it has been largely ignored by the "brain and mind (cognitive) sciences" in modern times. However, there are signs that this is beginning to change. In this article, we survey some issues about the nature of emotion, describe what is known about the neural basis of emotion, and consider some efforts that have been made to develop computer-based models of different aspects of emotion.

What Is Emotion?

The nature of emotion has been debated within psychology for the past one hundred years. The formal debate goes back to William James's famous question: Do we run from the bear because we are afraid, or are we afraid because we run? James suggested that we are afraid because we run. Subsequently, the psychological debate over emotion has centered on the question of what gives rise to the subjective states of awareness that we call feelings, or emotional experiences. Theories of emotional experience typically seek to account for how different emotional states come about and can be grouped into several broad categories: feedback, central, arousal, and cognitive theories (for review, see (LeDoux, 1996)). Though very different in some ways, each of these theories proposes that emotional experiences are the result of prior emotional processes. Feedback and arousal theories require that the brain detect emotionally significant events and produce responses appropriate to the stimulus; these responses then serve as a signal that determines the content of emotional experience. Central and cognitive appraisal theories, which are in some ways different levels of description of similar processes, assume that emotional experience is based on prior evaluations of situations; these evaluations then determine the content of experience. Interestingly, the evaluative processes that constitute central and appraisal theories are also implicitly necessary for the elicitation of the peripheral responses and arousal states of feedback and arousal theories.

The disparate theories of emotional experience thus all point to a common mechanism—an evaluative system that determines whether a given situation is potentially harmful or beneficial to the individual. Since these evaluations are the precursors to conscious emotional experiences they must, by definition, be unconscious processes. Such processes are the essence of the ignored half of James's question. That is, we run from a bear because our brain determines that bears are dangerous. Many emotional reactions are likely to be of this type: unconscious information processing of stimulus significance, with the experience of "emotion" (the subjective feeling of fear) coming after the fact.

Although the manner in which conscious experiences emerge from prior processing is poorly understood, progress has nevertheless been made in understanding how brain circuits process emotion. Just as vision researchers have achieved considerable understanding of the neural mechanisms underlying the processing of color while still knowing little about how color experience emerges from color processing (see COLOR Perception), it is possible to study how the brain processes the emotional significance of situations without first solving the problem of how those situations are experienced as conscious content.

The Neural Basis of Emotional Processing

Traditionally, emotion has been ascribed to the brain's limbic system, which is presumed to be an evolutionary old part of the brain involved in the survival of the individual and species (LeDoux, 2000). Some of the areas usually included in the limbic system are the hippocampal formation, septum, cingulate cortex, anterior thalamus, mammillary bodies, orbital frontal cortex, amygdala, hypothalamus, and certain parts of the basal ganglia. However, the limbic system anatomical concept and the limbic system theory of emotion are both problematic (LeDoux, 2000). The survival of the limbic system theory of emotion is due in large part to the fact that the amygdala, a small region in the temporal lobe, was included in the concept.

The amygdala has been consistently implicated in emotional functions (see (Damasio, 1999; LeDoux, 1996; Rolls, 1998), and various chapters in (Aggleton, 1992)). Lesions of this region interfere with both positive and negative emotional reactions. Moreover, unit-recording studies show that cells in the amygdala are sensitive to the rewarding and punishing features of stimuli and to the social implications of stimuli. Other limbic areas have been less consistently implicated in emotion, and when they have been implicated, it has been difficult to separate out the contribution of the region to emotion per se as opposed to some of the cognitive prerequisites of emotion. The amygdala therefore serves as an experimentally accessible entry point into the distributed network of brain regions that mediate complex emotional evaluations.

The contribution of the amygdala to emotion results in large part from its anatomical connectivity (LeDoux, 2000; Pitkanen, Savander, & LeDoux, 1997). The amygdala receives inputs from each of the major sensory systems and from higher-order association areas of the cortex. The sensory inputs arise from both the thalamic and cortical levels. These various inputs allow a variety of levels of information representation (from raw sensory features processed in the thalamus to whole objects processed in sensory cortex to complex scenes or contexts processed in the hippocampus) to impact on the amygdala and thereby activate emotional reactions. Most of these sensory inputs converge in the lateral nucleus of the amygdala, and the higher order information in the basal nucleus. These can be viewed as the sensory and cognitive gateways, respectively, into the amygdala's emotional functions. At the same time, the amygdala sends output projections to a variety of brainstem systems involved in controlling emotional responses, such as species-typical behavioral responses (including facial expressions and whole body responses such as freezing), autonomic nervous system responses, and endocrine responses. Most of these outputs originate from the central nucleus of the amygdala. Recent anatomical and physiological work has however shown that the amygdala consists of several interacting subnuclei that may have specific individual contribution to the overall emotional computation performed (see below). If the amygdala is consistently found to contribute to the evaluation of the emotional significance of a stimulus, are there systems that control the processing of the amygdala? Recent work suggest that the amygdaloid complex can be modulated by neurochemical systems such as serotonergic or dopaminergic, that are activated in relation to the overall behavioral state of the organism.

Much of the anatomical circuitry of emotion described above has been elucidated through studies of fear conditioning, a procedure whereby an emotionally neutral stimulus, such as a tone or light, is associated with an aversive event, such as a mild footshock (Davis, 1998; Fendt & Fanselow, 1999; LeDoux, 2000). After such pairings, the tone or light comes to elicit emotional reactions that are characteristically expressed when members of the species in question are threatened. While there are other procedures for studying emotion, none has been as successfully applied to the problem of identifying stimulus-response connections in emotion. The fear conditioning model is at this point particularly attractive since it has laid out pathways from the sensory input stage to the motor output stage of processing, showing how simple stimulus features, stimulus discriminations, and contexts control the expression of behavioral, autonomic, and endocrine responses in threatening situations.

Although many emotional response patterns are hard-wired in the brain's circuitry, the particular stimulus conditions that activate these are mostly learned by association through classical conditioning. The amygdala appears to contribute significantly to this aspect of learning and memory and may be a crucial site of synaptic plasticity in emotional learning (Fendt & Fanselow, 1999; LeDoux, 2000). This form of memory is quite different from what has come to be called *declarative memory*, the ability to consciously recall some experience from the past. Declarative memory, in contrast to *emotional memory*, crucially requires the hippocampus and related areas of the cortex. When we encounter some stimulus that in the past had aversive consequences, we recall the details of who we were with and where we were and even that it was a bad experience. However, in order to give the declarative memory an emotional flavor, it may be necessary for the stimulus, simultaneously and in parallel, to activate the emotional memory system of the amygdala. It is likely to be this dual activation of memory systems that gives our ongoing declarative memories their emotional coloration. Emotional memories are formed by the amygdala, in the same manner as declarative memories are formed in the hippocampus. The actual site of storage of emotional and declarative memories is still a matter of debate, but may involve distant cortical and subcortical areas in addition to the amygdala and hippocampus (Fanselow & LeDoux, 1999).

In the last several years, the basic findings regarding fear conditioning in animals have been confirmed and extended by studies of brain damaged patients and functional imaging studies. This work has shown that the human amygdala is also involved in fear learning and other emotional processes (for review see (Damasio, 1999; LeDoux & Phelps, 2000 ; Phelps & Anderson, 1997)).

At this point, we have mentioned "emotional experience" a number of times, and it may be worth speculating on just what an emotional experience is and how it might emerge. The emotion of fear will be used as an example. All animals, regardless of their stage of evolutionary development, must have the ability to detect and escape from or avoid danger. The widespread distribution of these behaviors in the animal kingdom makes it unlikely that the subjective experience of fear is at the heart of this ability. It may well be the case that subjective, consciously experienced fear is a mental state that occurs when the defense system of the brain (the system that detects threats and organizes appropriate responses) is activated, but only if that brain also has the capacity for consciousness. That is, by this reasoning, fear and other emotions reflect the representation of the activity of neural systems shaped by evolution and the responses they produce as conscious content. If this is true, then it is important that we focus our research efforts on these stimulus detection-and-response organizing systems, as these are the systems that generate the conscious content we call emotions. While emotional behaviors may be triggered by sensory inputs that bypass or pass through the neocortex, the experience of emotion is likely to involve the cortical representation of the emotional episode. Although our understanding of the cortical representation of emotion episodes (or other conscious experiences) is poor at present, considerable evidence suggests that working memory circuits involving the frontal lobe may play a key role (LeDoux, 2000).

Computational Models of Emotion

Using computers to understand emotions has always been a challenge. Popular beliefs define computing devices as inherently incapable of exhibiting and experiencing any emotions and, at present, no definite claims have been made that computers may be suitable for such a task. Nevertheless, consistent with the notion put forth in the introduction, computers are used as tools for modeling certain aspects of emotional processing.

Models of Emotional Learning and Memory

As proposed by most central theories, many emotional responses are hard-wired in brain circuitry. Nevertheless, in humans and animals, the environmental events that trigger these responses are often learned through experiences in which emotionally neutral stimuli come to be associated with emotionally charged stimuli. One important aspect of emotional processing, therefore, involves the manner in which the brain forms, stores, and uses associations between meaningless and meaningful stimuli.

(Grossberg & Schmajuk, 1987) developed a model of conditioned affective states based on the notion that conditioned reinforcement involves pairs of antagonistic neural processes, such as fear and relief. Their model suggests a mechanism by which neutral events are charged with a reinforcing value (either positive or negative) depending on the previous activity of the model. The simulated neural circuits are suggestive of the role of brain structures involved in the processing of certain emotions, such as the hippocampo-amygdaloid system (described as a zone of convergence of conditioned (CS) and unconditioned (US) stimulus pathways), the septum (described as a zone in which the opposition of the processes is represented), the hypothalamus, the nucleus of the solitary tract, and the reticular formation (described as zones of visceral and somatosensory inputs).

In an extension of the basic model, Ricart added a new neural center to the US pathway, the role of which is to prolong the neural representation of the US after its actual termination (Ricart, 1992). The amount and nature of the activity of this center is related to the "unexpectedness" of the stimuli and is analogous to the activity of the locus coeruleus (LC), which is known to be involved in attention.

Armony and co-workers have implemented another connectionist model of emotional learning and memory that, like the previous two conditioning models, also focuses on zones of convergence of US and CS pathways (Armony, Servan-Schreiber, Cohen, & LeDoux, 1997). In contrast to Grossberg and Schmajuk's model and Ricart's model, this model is anatomically constrained by the known data of the fear conditioning circuitry. It examines processing in two parallel sensory (CS) transmission pathways to the amygdala from the auditory thalamus and the auditory cortex in a learning situation involving an auditory CS paired with a footshock US. The model is initially trained using a modified Hebb-type learning rule and, under testing conditions, reproduces data related to frequency-specific changes of the receptive fields known to exist in the auditory thalamus and amygdala. The model predicted that lesions of the cortical auditory route would not affect the specificity of the behavioral response to a range of frequencies centered on the training (aversively meaningful) frequency. This prediction has been verified experimentally. Because cortical representations are subject to attentional focus, this modeling study, like the previous one, suggests a close link between the amygdala and the attentional system of the midbrain.

Recent anatomical studies coupled to in-vivo and in-vitro physiological experiments have provided invaluable data that can be used to build biophysically realistic computational models of amygdala circuits. Such models explore the interactions between converging thalamic and cortical inputs onto neurones in the lateral nucleus of the amygdala (Armony & LeDoux, 1997), as well as the role of local feedforward and feedback inhibition in stimulus processing (Li, Armony, & LeDoux, 1996).

Computational Models of Cognitive-Emotion Interactions and Appraisal

Researchers in experimental psychology, artificial intelligence (AI), and cognitive science have long recognized the mutual influences of emotions and cognition. However, these interactions are still not clearly formulated. We still do not have adequate theories defining each of these components of human mentation (emotion and cognition), much less a full understanding of how cognition and emotion might relate (see EMOTION-COGNITION INTERACTIONS).

As described above, most theories of emotion recognize the importance of evaluative or appraisal processes. Although there is considerable disagreement as to how these processes should best be viewed, most workers nevertheless see evaluative or appraisal processes as functioning by comparing sensed characteristics of the world to internal goals, standards, and attitude structures, deducing the emotional significance of the stimulus, guiding the expression of emotional behavior and other physiological responses, and influencing other modules pertaining to behavioral decisions.

In principle, it is possible to model appraisal processes using classical symbolic AI techniques (see (Picard, 1997) and (Chwelos & Oatley, 1994), for reviews). It is possible, for example, using a vector space approach, to find a plausible mapping between appraisal features (e.g., novelty, urgency, intrinsic pleasantness) and emotion categories (e.g., fear, joy, pride). Relying on a posterior verbal reports and a predefined set of emotions, one could then derive a limited set of appraisal criteria, sufficient for emotion prediction and differentiation. Other AI approaches, such as decision trees, pattern matching, and production rules (expert systems), are also possible, although each of these methods encounters theoretical difficulties. These types of systems, however, do not generally account for neurophysiological data.

One criticism often made of cognitive models of emotion is related to the complexity of processing involved and to the time they consequently require. From an AI point of view, the criticism has been addressed by introducing reactivity to "classical" cognitive models. Classical AI approaches assume that systems possess a well defined representation of their environment, state, actions, and goals. In contrast, reactive systems do not make such assumptions; they are mostly based on real-time, incomplete evaluations, their performance being based more on the properties of the evaluative mechanisms than on the quality and quantity of their internal representations (see (Lyons & Hendricks, 1992) for a review and examples).

It is interesting to note that, as we mentioned earlier, appraisal of sensory information might be one of the most prominent functions of the amygdala, placing this structure in a key position to actually perform the mapping of the emotional value of the stimuli. In this view, the relation between amygdala activity and emotion is a computational one (in the broad sense of the term) rather than a subjective one. The existence of multiple pathways to the amygdala from input processing systems of various levels of complexity (see above) provides a biological resolution to some of the concerns that have been raised about the importance of cognition in driving emotion. The involvement of cognition can be minimal or maximal, depending on the situation.

Models of Facial Expressions of Emotion

Of interest to feedback and arousal theories, the expression of emotion in the face is an important biological aspect of emotion that has significant implications for how emotion is communicated in social situations (Darwin, 1872). Face recognition and analysis of facial expression has only recently been an active field of research in the computer vision community (for review, see (Samal & Iyengar, 1992)). Face analysis can be computationally decomposed into three sub-problems: detecting the face in a scene; identifying the face; and analyzing its expression. At present, each of these tasks uses different features of the face, and different computational approaches. These approaches are based on psychophysical observations and are not yet explicitly based on neurophysiological data. However, a number of neurophysiological studies have been conducted (for review, see Rolls's chapter in (Aggleton, 1992)). These studies have shown cells selectively responsive to particular faces in areas of temporal neocortex and in the amygdala. More recently, other studies showed that there might be an influence of facial expressions on the actual neural correlates of the emotional states experienced, through modifications of blood flow characteristics (for review, see (Ekman, 1992)). Other approaches are more physico-mathematical, relying on image processing techniques. These implementations address exclusively the problem of emotional expression (and, possibly, communication of emotions) without relying on any theory of emotional experience (Bartlett, Hager, Ekman, & Sejnowski, 1999).

Conclusion

It is important to distinguish between emotional experiences and the underlying processes that lead to emotional experiences. One of the stumbling blocks to an adequate scientific approach to emotion has been the focus of the field on constructing theories of the subjective aspects of emotion. Studies of the neural basis of emotion and emotional learning have instead focused on how the brain detects and evaluates emotional stimuli and how, on the basis of such evaluations, emotional responses are produced. The amygdala was found to play a major role in the evaluation process. It is likely that the processing that underlies the expression of emotional responses also underlies emotional experiences, and that progress can be made by treating emotion as a function that allows the organism to respond in an adaptive manner to challenges in the environment rather than to a subjective state. While computational approaches to subjective experiences of the emotional or non-emotional kind are not likely to be easily achieved, computational approaches to emotional processing are both possible and practical. Although relatively few models currently exist, this situation is likely to change as researchers begin to realize the opportunities that are present in this too-long neglected area.

Road Map: Connectionist Psychology
Related Reading: Conditioning; Emotion-Cognition Interactions; Sparse Coding in the Primate Cortex

References

Aggleton, J. P. (Ed.). (1992). *The Amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction*. New York: Wiley-Liss.

Armony, J. L., & LeDoux, J. E. (1997). How the brain processes emotional information. *Ann N Y Acad Sci, 821*, 259-70.

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., & LeDoux, J. E. (1997). Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning. *Trends in Cognitive Sciences, 1*, 28-34.

Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology, 36*(2), 253-63.

Chwelos, G., & Oatley, K. (1994). Appraisal, Computational Models, and Scherer's Expert System. *Cognition and Emotion, 8*(3), 245-257.

Damasio, A. (1999). *The Feeling of What Happens : Body and Emotion in the Making of Consciousness*: Harcourt Brace.

Darwin, C. (1872). *The Expression of Emotions in Man and Animals*. London: Julian Friedman.

Davis, M. (1998). Anatomic and physiologic substrates of emotion in an animal model. *J Clin Neurophysiol, 15*(5), 378-87.

Ekman, P. (1992). Facial Expressions of Emotion: New Findings, New Questions. *Psychological Science, 3*(1), 34-38.

Fanselow, M. S., & LeDoux, J. E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron, 23*(2), 229-32.

Fendt, M., & Fanselow, M. S. (1999). The neuroanatomical and neurochemical basis of conditioned fear. *Neurosci Biobehav Rev, 23*(5), 743-60.

Grossberg, S., & Schmajuk, N. A. (1987). Neural Dynamics of Attentionally Modulated Pavlovian Conditioning: Conditioned Reinforcement, Inhibition, and Opponent Processing. *Psychobiology, 15*(3), 195-240.

LeDoux, J. (1996). *The Emotional Brain*. New York: Simon & Schuster.

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience, 23*, 155-184.

LeDoux, J. E., & Phelps, E. A. (2000). Emotional networks in the brain. In M. Lewis & J. Haviland (Eds.), *Handbook of Emotions* (pp. in press). New York: Guilford Press.

Li, X. F., Armony, J. L., & LeDoux, J. E. (1996). GABAA and GABAB receptors differentially regulate synaptic transmission in the auditory thalamo-amygdala pathway: an in vivo microiontophoretic study and a model. *Synapse, 24*(2), 115-24.

Lyons, D. M., & Hendricks, A. J. (1992). Planning, Reactive. In S. C. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence* (Second Edition ed., Vol. 2, pp. 1171-1181). New York: John Wiley & Sons.

Phelps, E. A., & Anderson, A. K. (1997). Emotional memory: what does the amygdala do? *Curr Biol, 7*(5), R311-4.

Picard, R. W. (1997). *Affective Computing*: M.I.T Press.

Pitkanen, A., Savander, V., & LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends in the Neurosciences, 20*(11), 517-523.

Ricart, R. (1992). Neuromodulatory Mechanisms in Neural Networks and Their Influence on Interstimulus Interval Effects in Pavlovian Conditioning. In D. S. Levine & S. J. Leven (Eds.), *Motivation, Emotion, and Goal Direction in Neural Networks* (pp. 115-165). Hove and London: Lawrence Erlbaum Associates.

Rolls, E. T. (1998). *The Brain and Emotion*: Oxford University Press.

Samal, A., & Iyengar, P. A. (1992). Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. *Pattern Recognition, 25*(1), 65-77.