No Inferiority Complex in the Study of Emotion Complexity: A Cognitive Neuroscience Computational Architecture of Emotion[1]

David Sander & Olivier Koenig
Laboratoire d'Etude des Mécanismes Cognitifs, Université Lyon 2[2]
& Institut des Sciences Cognitives, Bron

The aim of the present paper is to propose, on the basis of a survey of the relevant literature, an explicitly-described computational architecture of the emotion system. We first argue that cognitive scientists have the legitimacy to study emotions and that cognitive neuroscience concepts and methods are critical for the elaboration of such an architecture. Then, we propose some functional, computational and nervous system related principles that can constrain the elaboration of this architecture. Finally, this framework leads us to the description of the subsystems that constitute the postulated computational architecture.

Keywords: emotion, computational architecture, affect, computational analysis, functional architecture, cognitive neuroscience

Introduction

Since the 1950's, cognitive scientists have dramatically neglected emotions. At first sight, this disinterest can appear incongruous if one considers how salient and functional emotions are in our everyday life. At least two explanations of such a disinterest can be highlighted.

First, emotions were considered to be too complex. Therefore, cognitive scientists have developed what one may call an inferiority complex. It is striking that cognitive scientists felt having neither the tools nor the legiti-

---

[1] This paper was written thanks to a grant from the "Cognitique" Program from the French Ministry of Research (project "COG 58," entitled "Etude de l'architecture fonctionnelle des mécanismes émotionnels: psychologie cognitive, neuropsychologie, psychopathologie et imagerie cérébrale fonctionnelle," year: 2000, amount: 500 000 FF, duration: 2 years).

[2] Addresses for correspondence: David Sander or Olivier Koenig, Laboratoire d'Etude des Mécanismes Cognitifs, Université Lyon 2, 5 Avenue Pierre Mendès-France, 69676 Bron Cedex, France. E-mail : sander@isc.cnrs.fr or koenig@univ-lyon2.fr (Fax : +33 (0)4 78 77 43 51).

Cognitive Science Quarterly (2002) 2, 249-272

macy for investigating emotion. But an "emotional revolution" took place both in cognitive neuroscience (CN) and in artificial intelligence (AI). Underlying this revolution are, probably, the evolutionary perspective (see Tooby & Cosmides, 2000), the growing use of brain imaging techniques in CN during the 1990's (see Gazzaniga, 2000), and the emergence of affective computing (see Picard, 1997) and embodied perspective in AI (see Cañamero, 2001). The inferiority complex could disappear if this revolution – also promoted by parallel "social cognition" and "social functions" studies (see Adolphs, 2001) – were to be reinforced by a computational approach.

Second, this disinterest becomes obvious if one considers that the dominant mode of thinking, deeply marked by a scientifically-correct cartesianism, was to conceive the cognitive system as the "incarnation of reason." Therefore, emotions were traditionally considered as troubles of the (cognitive) mind. As an implication, until recently, emotion was not in the scope of traditional cognitive science. This dilemma that reveals an ambiguous relation between the concepts of emotion and cognition can be solved by arguing that emotion is part of cognition. Indeed, the emotion system can be seen as a particular cognitive system – here defined as a natural or artificial system that processes (not necessarily in a symbolic way) information that serves to acquire, organize or use knowledge.

In the following, we first expose how the computational approach of cognitive neuroscience can be applied to the study of emotions (Section 1). Then, we develop principles constraining the elaboration of a computational architecture of emotion (Section 2). Finally, we propose a computational architecture of the emotion system based on these principles (Section 3).

1. A computational approach for the cognitive neuroscience of emotions

1.1 The computational analysis as a cognitive neuroscience tool

Since its emergence, in the 1980's, CN has adopted an original position within the cognitive science framework by attempting to integrate complementary methods and concepts. As outlined in the following, CN can be described through its goals and methods.

The goals of cognitive neuroscience

The goals of CN are ambitious. The classical view is that *"Cognitive neuroscience studies are beginning the task of integrating questions of human cognition from neurons through behavior"* (Posner & DiGirolamo, 2000, p. 881). More precisely, our view is that a challenge for cognitive neuroscientists is to investigate the cognitive system by studying cognitive mechanisms and their interactions in a biologically plausible way in order to produce computational models. By "biological plausibility", we mean that designing a cognitive model is constrained by what is known about the nervous system. This latter statement implies that models that only take into account either functional or brain aspects are not cognitive neuroscience products. By referring to computational models, we argue that an efficient way to elaborate models is to perform computational analyses. One can define a computational analysis, in the framework of cognitive neuroscience, as *"a logical exercise aimed at determining what processing subsystems are necessary to produce a specific behavior, given specific input"* (Kosslyn & Koenig, 1995, p. 41). Computational and biological constraints imply that the goal of CN is not to build any model that can efficiently simulate a given set of behaviors but, instead, to design cognitive models that are consistent with behavior, simulation and brain data. We believe that adopting such a perspective offers at least two major advantages.

First, a computationally and biologically constrained model allows one to test predictions regarding the normal cognitive system. So, theoretically, such a model becomes a tool that can lead to new insights about natural cognition. Ideally, if the computational model is explicit enough, it can help designing computer models that simulate the operations of these subsystems. Such models can be implemented using, for instance, artificial neural networks or autonomous agents. For example, as explained by Cañamero (2001, p. 527), we believe that *"autonomous agents provide an excellent testbed to study the nature and adaptative value of emotions, with a synthetic value that nicely complements analytic studies of natural emotions."*

Second, natural-like computational models can help a pragmatic designer to build more efficient artificial systems. Indeed, natural organisms solve problems in a way that can be a powerful source for inspiration. For instance, this is what happened for the invention of artificial neural networks. Nevertheless, let us notice that designing systems that are very efficient without being consistent with CN data may be, by itself, a fascinating field of artificial intelligence. Evolution of living species is still on its way and there are many alternative procedures for an artificial system to perform better than we do. For example, it would not have been efficient to wait until we know how the human brain computes numbers before designing the first calculator. From this perspective, it was proposed that a way to resolve the "emotion complexity" problem in AI would be to create artefacts that express recognizable emotions, no matter how they are built (for discussion, see Cañamero, 2001).

The methods of cognitive neuroscience

CN provides a multimethodological answer to the question of how emotional models inspired by natural cognition models can be built. The interdisciplinary approach of CN arises mainly, but not only, from the following fields. First, cognitive psychology is the leading field for testing the functional organization of normal cognitive mechanisms in healthy subjects. Second, cognitive neuropsychology allows for testing hypotheses about the normal cognitive system by designing experiments involving patients with focal brain lesions. Third, psychopathology can be seen as the continuity of neuropsychology by considering patients suffering from mental illness as having the same epistemological value as neuropsychological patients (e.g., Frith, 1992). Fourth, brain imaging techniques play a central role in CN, revealing which parts of the brain are involved in a particular task. Fifth, connectionism serves CN both as a theoretical paradigm and as an experimental field.

1.2 Functional architectures for emotions

Adopting an evolutionary perspective, explaining what the brain is for should be done in computational terms *"because that is the only language that can capture or express the functions that neural properties were naturally selected to embody"* (Tooby & Cosmides, 2000, p1168). Hence, the main goal of the computational analysis of the emotion system is to use this language in order to build a functional architecture. Let us notice that even if the functional approach to emotion is largely shared, most functionalists do not use the cognitive neuroscience approach when studying the emotion system.

1.3 Definitional issues

Unsolved difficulties are encountered both when attempting to separate emotion from other phenomena and when searching for a definition of emotion *per se*. Emotion is often discussed in a framework including other affective phenomena, such as affect, feeling, motivation, passion, mood, affective style, affective reactivity, or drives. Some of these concepts suffer less than others from their "folk" meaning because they are scientific concepts (e.g., *Affective Style*, see Davidson, 1992). However, we agree with Sloman (2001) who points to the ambiguous meanings of such terms as "emotion", "feeling", and "mood". For example, because Rolls (1999) included thirst or sexual behavior as emotions in his book *"The Brain and Emotion"*, Phillips (1999) proposed that it might have been more appropriate to entitle this book *"The Brain and Motivation."* Hence, it is a scientific challenge to identify some of these phenomena, and in particular emotion.

Emotion as the studied object

A consensual definition of emotion has never been found, neither in CN nor in any other science (see Strongman, 1996). For example, Kleinginna and Kleinginna (1981) listed ninety-two definitions of emotion. It is intuitive that researchers who belong to distinct disciplines may disagree about a definition. For example, two opposite positions are the phenomenological view arguing that emotions are particular modes of consciousness (e.g., Sartre, 1938) and the behavioral view that defines emotions as *"states elicited by rewards and punishers"* (Rolls, 1999, p.60). Moreover, the definitional disagreement can also take place within a discipline such as neuroscience. For example, according to Damasio (1998, p.84), *"the term emotion should be rightfully used to designate a collection of responses triggered from parts of the brain to the body, and from parts of the brain to other parts of the brain."* On the other hand, LeDoux (1994, p. 291) wrote that *"in my view, emotions are affectively charged, subjectively experienced states of awareness. Emotions, in other words, are conscious states."* This divergence has important implications; for example, according to LeDoux emotions are conscious phenomena, whereas according to Damasio they are not. However, LeDoux calls *"emotion"* what Damasio calls *"feeling."* Thus, there appears to be more a purely terminological confusion than a scientific discord.

Instead of proposing a concise definition of emotion, we intend to participate to the characterization of its functions, its computational architecture and its implementation in the brain. Hence, in the following, we propose an operational exercise consisting in elaborating a set of principles which a computational model of emotion should ideally rely on.

2. Principles for a computational model of emotion

Any CN computational model of emotion should follow two main kinds of principles. First, it should follow principles that must be fulfilled by any scientific theory, such as refutability. We are not going to develop these principles in the present paper because they are general (see Popper, 1968). Second, it should follow specific principles constrained by the interdisciplinary approach to emotions. These principles are of critical interest for our purpose. Therefore, we propose a list of principles divided into function, computation and nervous system related principles.

2.1 Functional principles (FP). The 3 Es: evaluation, expression and experience

Evaluation

The mechanisms involved in emotional evaluation are those that allow us to extract an emotional value from external stimuli. The emotion literature identifies several emotional functions and we propose six principles that seem critical for the evaluative component of emotions.

(FP1) To detect relevant stimuli. From an evolutionary perspective, a major function of emotion is to signal relevant environmental events in order to mobilize the organism (see Frijda, 1994; Ekman, 1999).

(FP2) To apply the precaution principle. To be optimal, the detection of relevant stimuli must use minimal information and not delay the response. In some cases, the stimulus does not need to be identified in order to trigger an emotional response; to use a metaphor from Frijda (1994), if risks are potentially high "you shoot before you ask questions" (see also the computational principles section). This precaution principle can lead to positive errors because non emotional stimuli can be mistaken for emotional elicitors (e.g., mistaking a piece of deadwood for a snake when walking in a wood induces a fear reaction; see LeDoux, 1996).

(FP3) To signal the emotional value to other cognitive systems. The detection of relevant stimuli allows the organism to provide on-line information about their emotional value to other cognitive systems such as memory and attention. The minimal signal provided should be in categorical terms such as "the stimulus is good or bad" for the organism. Moreover, as the emotional value is not reduced to the hedonic dimension (see Cacioppo & Gardner, 1999), the signal should also provide information relative to the emotional intensity. For example, the more intense the stimulus, the faster the emotional reaction (see Scherer, 1994) and the better the stimulus is remembered (Canli, Zhao, Brewer, Gabrieli & Cahill, 2000).

(FP4) To prioritize some goals and processes. When considering the relations between the concepts of "goal" and "emotion", one can distinguish between at least three topics. First, emotion can be seen as a goal *per se*. Indeed, particular emotional experiences can be actively pursued or avoided. For example, when visiting the space mountain attraction in Disney parks, one's goal is *only* to experience a specific emotion. Second, emotion is involved in the selection of a goal. When one is prioritizing a particular goal among alternatives, emotions can guide the selection process. For example, consider the situation of a wild bear escaping from his cage in a zoo. If one is directly confronted with the dangerous animal, fleeing from the bear will have higher priority than warning the guard, whereas if one is far away from the bear, the highest priority will be to warn the guard. Third, a major function of emotion is to prepare and support the achievement of a goal. For example, the emotion felt by a pianist playing live may help him to concentrate and achieve his goal to perform as well as possible.

(FP5) To interrupt non-relevant ongoing processes. A corollary of the prioritizing function is the interruption of any ongoing process that is not consistent with the evaluative process. For example, if one opens a closet to take one's coat and is suddenly facing a hairy spider, the on-going process to get dressed is interrupted.

(FP6) To permit emotional learning. The fear conditioning paradigm has shown that, under certain circumstances, one is able to attribute an emotional value to a previously neutral stimulus if it has been associated to an emotional unconditional stimulus (see Bechara et al., 1995; LeDoux, 1996). The emotional attribution to the conditional stimulus is only possible because we are able to evaluate, at the same time, the unconditional stimulus as emotional. Thus, evaluation can lead to confusion. An example can be found in Clore (1994, p107) writing that *"one of the oldest advertising strategies* *is to encourage male consumers to confuse their interest in an alluring female model with their interest in the product she is associated with."*

Expression

We consider here the process underlying the production and not the perception aspect of emotional expression. The expression aspect of emotion covers observable changes in the body and the behavior that correlate with experiencing an emotion. Such parameters include autonomic changes (e.g., heart rate, electrodermal response), motor changes (e.g., facial expression, posture), endocrine variations and prosody. We propose three main principles that seem to depict the functional value of emotional expression.

(FP7) To communicate an emotional state. Darwin (1872) was the first to propose that being able to express emotions is an evolutionary advantage. Indeed, direct expression, for example through the face, is the fastest unambiguous way to signal our emotion to other individuals. An example, proposed by Scherer (1994), can be chosen from social interaction: *"the facial reaction to an ironic statement will reveal immediately whether the utterance was perceived as a joke or an insult."*

(FP8) To communicate intentions. A corollary of the previous function is that expressing an emotion allows observers, who have a theory of mind, to attribute specific intentions. The theory of mind is the ability to ascribe a mental state to oneself and to others (Perner & Lang, 1999). For example, if you express anger, someone may ascribe you the intention to fight.

(FP9) To modulate/elicit emotional experience. The first author to systematize the role of emotional expression in one's own emotional experience was James (1884). According to this author, the perception of body changes that follow the processing of an emotional stimulus *is* the emotion. The role of body changes in evaluation and experience of emotion was discussed further by Damasio (1994) in his somatic markers' hypothesis.

Experience

We define an emotional experience as a feeling: The subjective conscious state of having an emotion. Because experience is caused both by evaluation and expression, it is uncertain if the four following functions are specific to experience or if they derive from evaluation and expression.

(FP10) To modulate explicit memory. Emotion and memory have strong connections. For example, emotional stimuli modulate explicit long-term memory storage (e.g., Cahill, 2000) and emotional arousal modulates explicit memory during both memory encoding and consolidation (Hamann, 2001).

(FP11) To modulate our judgements and decision-making. Experiencing an emotion has an effect on how one evaluates and judges stimuli (Clore, 1994). Moreover, experiencing an emotion can trigger unconscious processes that influence decision making (see Bechara, Damasio, Tranel & Damasio, 1997).

(FP12) To serve our theory of mind. Being able to experience emotions can serve our theory of mind. Indeed, knowing what it is like to have an emotion allows us, maybe through mental simulation (see Gallese & Goldman, 1999), to know what it is like for someone else to have an emotion. A primary observable information on which one relies in order to attribute the mental state of "having an emotion" is the facial expression. For example, this ability can lead to social interactions if someone recognizes sadness on a face.

(FP13) To attribute the cause of emotions to specific elicitors. If the elicitor of an emotion is not specified, the situation that conveys this emotion cannot be explicitly pursued or avoided. So, being conscious of having an emotion in a particular situation is an advantage because it allows one to store the elicitor event in explicit long term memory. This function is consistent with the affect-as-information model proposing that *"a primary function of emotion is to provide information about how a situation has been appraised. This information is conveyed internally by emotional experience"* Clore (1994, p. 104). The identification of eliciting situations constrains the generation of goals and plans. For example, someone who is conscious of being afraid in planes can take the train because he has identified "being in a plane" as a potential elicitor.

2.2 Computational principles (CP)

We propose some computational principles (CP) in order to constrain the architecture. A distinction is made between input, intermediate and output related constraints.

Input

(CP1) Exogenous and endogenous inputs. Two kinds of information enter the emotion system of one's organism: (1) information relative to external stimuli provided through exteroceptive perception and (2) information relative to one's internal somatic state provided through interoceptive perception (e.g., Bechara et al., 1995). Both types of information are complementary and probably necessary to elicit emotional experience. For example, Adolphs, Damasio, Tranel, Cooper, and Damasio (2000) showed, by testing brain damaged patients, that recognizing emotions from visually presented facial expressions requires right somatosensory-related cortices.

(CP2) Mental representations as inputs. Neither external nor internal information need to be actually present in order to elicit emotions. Indeed, the activation by mental imagery of those representations that are activated in the presence of stimuli (e.g., visual, auditory... but also interoceptive stimuli) are enough to elicit emotions. For example, because internal somatic states are represented in somatosensory-related cortices, it is possible for the brain to activate information relative to the internal somatic state via *'as-if-body-loops'* (see Damasio, 1994) simulating internal states in these cortices.

(CP3) Variations in eliciting potential of exogenous stimuli. Growing evidence suggest a selectivity of input entering the emotion system indicating that the latter would be genetically prepared to process certain classes of stimuli. It was proposed that some inputs elicit emotions because they are evolutionarily salient (e.g., Ekman, 1999). For example, it was suggested that fear reactions are preferentially activated by fear-relevant and evolutionary-salient stimuli (Öhman & Mineka, 2001).

Intermediate interface

(CP4) Polarity-dependant mechanisms. We propose that different subsystems are used to compute positive and negative emotional stimuli. This hypothesis is motivated by the fact that the expected behavior elicited by positive and negative stimuli is so different (e.g., to approach a positive stimulus and to escape from a negative one) that different systems subserve these behaviors. This position is consistent with the fact that fine motor control processes, needed to approach a stimulus, and fast motor processes used to escape require distinct systems.

One way to demonstrate that positive and negative emotional stimuli are processed in different subsystems is to show that different parts of the brain (e.g., left *versus* right hemisphere) process each kind of information (see the "brain-related principles" section of the present paper; see also Sander & Koenig, submitted). Additional evidence comes from the study of the amygdala. Although a growing number of studies show a clear involvement of the amygdala for the processing of positive stimuli (e.g., Hamann & Hui, 2002), evidence from animal research (see LeDoux, 1996), patient data (see Adolphs et al., 1999) and brain imaging studies (e.g., Whalen et al., 1998) suggest that this structure is more involved in the processing of negative stimuli that in the processing of positive stimuli. An example of a polarity-dependant subsystem is the "fear module" proposed by Öhman and Mineka (2001), which relies mainly on the amygdala.

(CP5) Automatic evaluative mechanisms. Evidence for automatic emotional processing has emerged from many research areas and paradigms such as the emotional priming paradigm, emotional conditioning, brain imaging and patient studies. Fazio, Sanbonmatsu, Powell and Kardes (1986) were the first to demonstrate an emotional congruency effect employing a priming method. Using a Stimulus Onset Asynchrony (SOA) of 300 ms, these authors showed that the time necessary for evaluating a target word was faster if the prime word was emotionally congruent. This effect has been replicated in many studies (e.g., Hermans, de Houwer & Eelen, 1994), provided that the SOA value was relatively short, such as 300 ms. The fact that the effect does not appear with a longer SOA (e.g., 1000 ms) supports the hypothesis of an automatic evaluative mechanism (see Klauer & Musch, 2002). In addition, Öhman and Soares (1998), using fear-relevant masked stimuli in an emotional conditioning experiment, observed that participants were able to extract emotional information, although they were unable to recognize stimuli. Furthermore, brain imaging studies have proven to be useful in the investigation of automatic emotional processing. Indeed, the amygdala has been found activated during unconscious processing of masked emotional stimuli (e.g., Morris, Buchel & Dolan, 2001; Whalen et al., 1998). Moreover, testing a blindsight patient, de Gelder, Vroomen, Pourtois, and Weiskrantz (1999) showed that recognition of emotional stimuli that were "not seen" was possible. Additional support for the hypothesis of automatic emotional evaluation was found by Vuilleumier and Schwartz (2001) who tested neglect patients and showed that negative stimuli were able to capture attention in a neglected field.

(CP6) Quick onset and a brief duration of emotions. A major difference between emotion and other affective phenomena is that the former has a relatively quick onset and brief duration (see Scherer, 2000). Hence, the dynamical interaction between the subsystems that gives rise to emotions takes place relatively quickly and does not last very long.

(CP7) Appraisal. The fact that emotions have a quick onset and a brief duration does not mean that endogenous and exogenous inputs cannot be appraised. Appraisal can be defined as a subjective evaluation of the meaning and consequences of an event, given a specific context and given one's own plan or goals. According to the component process model of emotion (see Scherer, in press), emotional elicitation and differentiation is the result of a sequential evaluation according to appraisal objectives (e.g., relevance, implications, coping potential, and normative significance).

Output

(CP8) Approach *versus* withdrawal. Davidson and Irwin (1999) proposed that an approach system facilitates appetitive behavior and generates some approach-related positive affects. These kinds of affects would be generated in the context of moving toward a goal. These authors also postulate a second system that facilitates the withdrawal of an individual and generates some withdrawal-related negative affects. Such a conception was also developed by Hobbes (1651, p.119) writing that *"This Endeavour, when it is toward something which causes it, is called Appetitive (...) And when the Endeavour is fromward something, it is generally called Aversion. These words Appetite, and Aversion (...) signify the motions, one of approaching, the other of retiring."*

(CP9) Emotional Expression. Main outputs of the emotion system are the expressive components. As described earlier, such components include autonomic changes (e.g., heart rate, electrodermal response), motor changes (e.g., facial expression, posture), endocrine variations and prosody.

(CP10) Non ambiguous expressive outputs of basic emotions. It was proposed that particular emotions, called basic emotions, are distinguishable one from the other and universally recognized (see Ekman, 1999). Ekman

(1999) proposed the following list of basic emotions: amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness/distress, satisfaction, sensory pleasure, and shame. Although not enough data is available to show that all these emotions are clearly distinguishable from another, evidence from cross-cultural research on facial expression supports the idea of universality in spontaneous expressions (see Ekman, 1999). The concept of basic emotions was also developed by Descartes (1631 Art. 69) who distinguished between six primitive emotions: admiration, love, hatred, desire, joy and sadness.

(CP11) Outputs as inputs. Emotion outputs are included in the set of inputs that enter the emotion system. Indeed, as already shown, it was shown that emotional expression modulates and elicits emotional experience.

2.3 Nervous-system principles

In this section, we propose some principles based on studies of the human nervous system in order to constrain the architecture. As shown earlier, the main argument is that a way to dissociate some subsystems is to show that they are implemented in different parts of the brain. A distinction is made between brain-based and somatic-based principles.

Brain-based principles (BP)

(BP1) Hemispheric asymmetry. To the question *"Why should there be some lateralization of emotional processing in human?"* (see Rolls, 1999), one could answer that if it is not necessary for an emotional function to be represented in both hemispheres due to the topology of the body, it is functionally relevant to place neurons that share the same function close together instead of distributing them in both hemispheres. Consistent with this logic, there is growing experimental literature reporting that some emotional processes are lateralized within the human brain. At least six hypotheses can be drawn from the literature related to the brain asymmetry of emotional processing.

1. "Right Hemisphere hypothesis"; it is mostly based on patient data and argues that there is a general right hemisphere advantage (RH-A) for the processing of emotional information (see Borod & Madigan, 2000).
2. "Valence hypothesis"; it argues that there is a center for positive feelings in the left hemisphere (LH) and a center for negative feelings in the RH. This hypothesis is based on clinical observations that patients with LH damage or inactivation showed a depressive-catastrophic reaction, whereas patients with RH damage or inactivation showed an euphoric reaction (for discussion, see Gainotti, 2000).
3. "Levels of Processing hypothesis"; it was originally based on patient data and is an alternative interpretation of the observations that led to the "valence hypothesis." This hypothesis postulates a RH-A for the

automatic processing of emotional schemata and a LH-A for emotional conceptualization and control (see Gainotti, 2000).

4. "Social versus Primary Emotions hypothesis"; this is based on lesion data and Wada tests and suggests that social emotions are modulated by the LH, whereas primary emotions are modulated by the RH (Ross et al., 1994).

5. "Prefrontal hypothesis"; it postulates that the left prefrontal cortex is more involved in approach-related emotional experiences and that the right prefrontal cortex is more involved in withdrawal-related emotional experiences (Davidson & Irwin, 1999). This hypothesis that highlights the approach-withdrawal dimension is close to the "Valence hypothesis" that highlights the positive-negative dimension. Evidence for a differential involvement of the left and right prefrontal cortex along the approach-withdrawal dimension has been found using electrophysiological (Davidson, 1995) and regional glucose metabolism measures (Sutton, Davidson, Donzella , Irwin & Dottl, 1997).

6. The "Revisited Valence hypothesis"; this stands for a corpus of results showing that some structures of the LH are more involved for negative emotions and, to a less extend, that some structures of the RH are more involved for positive emotions; so challenging the "Valence hypothesis." Considering hemispheric asymmetry, many results from brain imaging studies do not fit with the valence hypothesis (for a review concerning emotional experience, see Canli, 1999). Indeed, some studies found left, but not right insula activation during negative emotional experience (George, Ketter, Parekh, Herscovitch & Post, 1996) or during emotional processing of negative stimuli (Irwin et al., 1996; Morris, Öhman & Dolan, 1998, 1999). Furthermore, some studies found left, but not right, amygdala activation during negative emotional experience (Ketter et al., 1996), during emotional processing of negative stimuli (Breiter et al., 1996; Blair, Morris, Frith, Perret & Dolan, 1999; Morris et al., 1996, 1998; Sander et al., in preparation; Taylor et al., 1998), during implicit fear conditioning (Morris et al., 2001), during crossmodal binding of fear (Dolan, Morris & de Gelder, 2001), and during instructed fear (Phelps et al., 2001). Moreover, other structures than the amygdala and the insula have been found more activated in the LH during the processing of negative words (Maddock & Buonocore, 1997; Crosson et al., 1999), of aversive olfactory stimuli (Zald & Pardo, 1997), and of visual stimuli (Kosslyn et al., 1996). Consistent with brain imaging results presented above, some neuropsychological data also support this hypothesis. Indeed, it has been reported that lesions of the LH have a greater effect on the recognition of negative emotional vocalization (Pell, 1998). Moreover, testing patients with unilateral amygdala resection, Anderson and

Phelps (2001) suggested that the enhanced perception of aversive words depends on the left, but not right, amygdala.

Reviewing these hypotheses shows that experimental results relative to hemispheric lateralization of emotional processing seem to diverge in a not already understood way. But, to our view, these divergences may come from the fact that different hypotheses focus on different subsystems, and that different subsystems may be lateralized in different ways. Therefore, it is of interest to notice that inhibitory connections from prefrontal structures to the amygdala have been postulated on the basis of animal research and brain imaging studies. For example, conducting an fMRI study, Hariri, Bookheimer and Mazziotta (2000) showed that labeling some negative emotional expressions was associated with a diminished response in the amygdala correlating with a simultaneous increase of the response in the right prefrontal cortex. Moreover, Maratos, Dolan, Morris, Henson and Rugg (2001) showed that recognition of words presented in emotionally negative relative to neutral contexts was associated with enhanced activity both in *right* dorsolateral prefrontal cortex and *left* amygdala. To our view, these results do support the idea of a privileged relation between two structures from different hemispheres that are both involved in negative emotion, but at different levels: the right prefrontal cortex and the left amygdala.

(BP2) Brain connectivity and convergence zones. Brain structures involved in emotional processing form a highly interconnected network. The amygdala is connected to prefrontal, cingulate, temporal and occipital cortices, and to the hypothalamus and other subcortical structures (Amaral, Price, Pitkänen & Carmichael, 1992). The ventro-medial cortices are important for the linkage of exogenous and endogenous stimuli, because they contain convergence zones that presumably represent conjunctions of activity in networks coding for these stimuli (Damasio, 1994).

(BP3) The subcortical pathway. Morris et al. (1999) showed that regions of the pulvinar and superior colliculus covaried positively with the right amygdala during masked visual presentations of conditioned faces. These results led them to the conclusion that emotional value of visual stimuli could be detected and processed without conscious awareness by a colliculo-pulvinar-amygdala pathway. Results from de Gelder et al. (1999) showing that recognition of emotional stimuli is possible in a blindsight patient indicate that this pathway seems a good candidate for visual automatic emotional evaluation (see also Morris, de Gelder, Weiskrantz & Dolan, 2001). It was also shown that the amygdala is involved in fear conditioning in rats (LeDoux, 1996) and humans (Buchel & Dolan, 2000). Experiments of auditory fear conditioning in rats showed the existence of a subcortical pathway going directly from the auditory thalamus to the amygdala bypassing the cortex. The importance of such a pathway was confirmed by an anatomically-constrained connectionist model of the fear network (Armony, Servan-Schreiber, Cohen & LeDoux, 1997).

(BP4) A somatosensory map. The somatosensory cortices of the RH are the most complete map of the body state that the brain can use (Damasio, 1994). These structures that are critical for touch, temperature perception, pain sensation and visceral state, include the insular cortex, the primary (S1) and secondary (S2) somatosensory cortices. The anterior insula is a key structure because it receives inputs both from S1 and from the viscera.

(BP5) The anterior cingulate. The anterior cingulate is broadly connected to structures that we have already described as critical for emotional processing. Such structures include the anterior insula, the orbitofrontal cortex, the amygdala, and the hypothalamus (see Devinsky, Morrell & Vogt, 1995). Bush, Luu and Posner (2000), described two distinct parts of the anterior cingulate. One is involved in attentional processes, whereas the other is involved in assessing the salience of emotional and motivational information and in the regulation of emotional response.

Somatic-based principles (SP)

(SP1) Internal-state patterns are informative. An important contribution of the neuroscience of emotion was to show that the emotion system processes information derived from the body. When one is experiencing an emotion, one's internal state changes and this internal state pattern is information that is processed by the brain (e.g., Critchley, Mathias & Dolan, 2001).

(SP2) Internal-state patterns are distinctive. It was argued that the information provided by the internal-state patterns allows us to distinguish among different kinds of emotions (see Ekman, 1999). For example, although this question is still a matter of debate, distinctive internal-state patterns were found for most basic emotions (Ekman, Levenson & Friesen, 1983; Vernet-Maury, Alaoui-Ismaili, Dittmar, Delhomme & Chanel, 1999).

3. From principles to computational architecture.

Let us first mention that our aim was not to implement the principles we described into an artificial system. Rather, our aim was to revisit the computational architecture of emotion proposed in 1995 by Kosslyn and one of the present authors (Kosslyn & Koenig, 1995) in the light of the above-described principles. The second aim was to further develop this model by specifying some aspects of its architecture. More specifically, our contribution consisted in: (1) verifying whether the general framework of the model is consistent with the most recent literature; (2) decomposing two of the most important subsystems of this architecture, namely, the internal-state pattern activation subsystem and the emotion-instructions-generation subsystem, into finer-grained subsystems; and (3) delineating the connections and interactions between subsystems more precisely. Our approach, which we intended to further explain the proposed cognitive architecture, is a necessary step toward implementation. The architecture we discuss is illustrated in Figure 1. It is made of the following subsystems:

Somatosensory buffer. This subsystem is a short-term perceptive memory that represents, as a somatic map, the actual state of one's body. Its function is to structure endogenous inputs into perceptual units that constitute inputs for the somatotopic mapping subsystem, the internal-state preprocessing subsystem and the internal-state pattern activation subsystem. We propose that the somatosensory strip implements the somentosensory buffer (see BP4). This proposal is analogous to the one of retinotopic visual areas implementing the visual buffer involved in visual processing.

Somatotopic mapping subsystem. The function of the somatotopic mapping subsystem is to specify the body localization where an internal state change has occurred on the basis of the information provided by the somatosensory buffer. The information relative to the localization of the sources of particular bodily sensation is sent to associative memory. We propose that this subsystem is implemented in the posterior parietal part of RH (see BP4).

Internal-state preprocessing subsystem. We refer to the internal state, modified in particular by emotional expression (see FP9), as a set of information constituted of autonomic, endocrine, immune, muscular (e.g., facial) and prosodic responses. These pieces of information represent on-line cues processed by the emotion system (see CP1, SP1). The function of the internal-state preprocessing subsystem is to extract distinctive information about one's internal state. This distinctive information then feed the internal-state pattern activation subsystem. We propose that the internal-state preprocessing subsystem is implemented in the anterior insula (BP4).

Internal-state pattern activation subsystem. The preprocessing subsystem processes information that can be used for recognition, but the subsystem that stores representations of previous internal-state patterns in long-term memory is called internal-state pattern activation subsystem. It is postulated that patterns stored in this subsystem are distinctive and necessary for the recognition of one's body state (see SP1, SP2). We propose a division of this subsystem into a *positive* and a *negative internal-state pattern activation subsystem* (see CP4, BP1). The former stores internal-state representations that are activated when positive emotions are recognized; the latter stores internal-state representations that are activated when negative emotions are recognized. Output of both subsystems is a code that potentially activates representations in associative memory. We propose that this subsystem is implemented in the somatosensory cortices (BP4).

Stimulus-response connection subsystem. This subsystem implements "processing reflexes": a given stimulus elicits a given response. On the one hand, the striatum has been shown to be critical in the acquisition of learned habits and conditioned responses (see Mishkin, Malamut & Bachevalier, 1984). On the other hand, the amygdala has been shown to be critical in emotional conditioning (see FP6, BP3). This is why we postulate that the stimulus-response connection subsystem receives inputs from both the internal-state pattern activation subsystem and from pattern activation subsystems of other modalities. Moreover, we argue that this subsystem underlies innate emotional responses to evolutionary salient stimuli (e.g., snakes, spiders; see CP5, BP3). Thus, we propose that this subsystem is implemented in the striatum and the amygdala (see BP3).

Associative memory. The main function of associative memory is to identify a stimulus (see Kosslyn & Koenig, 1995). It stores amodal representations (i.e., representations that are not modality-specific and can be activated through any perceptual modality). The stimulus is identified when the input closely matches the features of a stored object. For example, the associative memory allows one to name the object "bear" if one sees or hears a bear. In addition, the associative memory allows to learn new associations (see FP10). In the case of emotional processing, the associative memory allows one to process an internal-state pattern in a particular *context* and to identify one's own emotion. The context is created by information about the situation provided by other perceptual modalities as well as ongoing plans and goals (see FP4, CP7, FP10-13). Furthermore, top-down information constitutes a signal from the associative memory to the internal-state pattern activation subsystem reflecting the fact that internal-state pattern can be activated in the absence of external stimuli (see CP2). Associative memory can be the locus of associations that point toward exteroceptive and interoceptive representations located in different pattern activation subsystems such as the visual, auditory, and internal-state pattern activation subsystems. These associations signal which regions were involved in simultaneous processing of internal and external stimuli (on the basis of inputs from the internal-state pattern activation subsystem and the pattern activation subsystems of exteroceptive modalities). These associations allow the reactivation of internal-state patterns on the basis of inputs from the subsystems involved in exteroceptive processing. We propose that these associations are implemented in the ventromedial prefrontal cortex (see BP2).

Emotion-instructions-generation subsystem. The function of the emotion-instructions-generation subsystem is to generate an initial specification of the brain/body profile that is appropriate when facing an emotional situation on the basis of information provided by associative memory. We propose a division of this subsystem into an *approach-related* and a *withdrawal-related emotion-instructions-generation subsystem* (see CP4, CP8, BP1). Perhaps the most critical aspect of the emotional reaction when an individual is facing a situation in which approach or withdrawal is necessary is the generation of emotional instructions. In an "approach situation", the adapted behavior is likely to be subtle and controlled, whereas in a "withdrawal situation", it is likely to be fast and less controlled. Therefore, emotional instructions are probably very different in each case and are likely to be generated by distinct subsystems. We propose that these subsystems are implemented, respectively, in the left and the right dorsolateral prefrontal cortex. In addition, the anterior cingulate which is involved in the regulation of the emotional response, may modulate the functions of these subsystems (see BP1, BP5).

Emotion-execution subsystems. The instructions generated by the emotion-instructions-generation subsystems must activate an "effector" subsystem in order to modulate the internal state. An emotion-execution subsystem is postulated to operate as an effector that converts the instructions into an actual modification of the internal state (see FP9, CP6, CP11). The output of this subsystem constitutes an input to the internal-state preprocessing subsystem and to the somatosensory buffer so that on-line information about the internal state can be processed. We propose that this subsystem is implemented in the amygdala, the hypothalamus and brainstem nuclei.

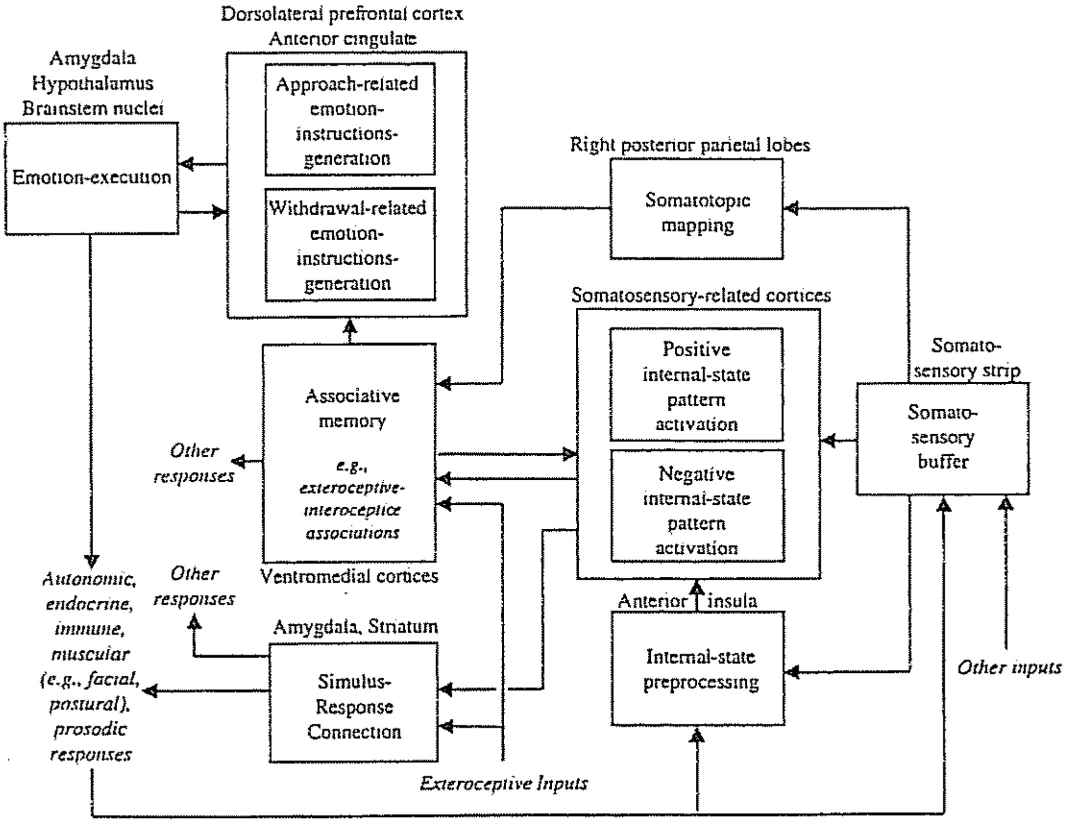

Figure 1. A proposal for a computational architecture of the emotion system inspired from Kosslyn and Koenig (1995)

An illustration

In order to provide an illustration of the data flow, inputs and outputs in the computational architecture, let us consider the following situation: an individual is walking in the wood. A particular configuration of the environment is processed by his exteroceptive systems (these systems, e.g., visual, are not represented in Figure 1). At the same time, his brain computes online information relative to his internal state. While walking in the wood, the evaluation of an emotional situation can be achieved by means of three processes.

The first process can lead to an emotional reaction in an automatic way, that is, without a necessary identification of the external stimulus. Suppose the individual confronts a snake. We assume that the following process will be the fastest to lead to an emotional reaction. Exteroceptive representations (e.g., a visual representation such as "the form of a snake" stored in the visual pattern activation subsystem) go through the stimulus-response connection subsystem and produce an emotional expression.

Let us now suppose that the individual, after a short walk, is asked to make a decision between keeping on walking or riding a horse. A second process may intervene here. Among the cognitive processes involved in the evaluation of such a situation, we argue that the exteroceptive-interoceptive representations stored in associative memory will play a decisive role. If the individual has already experienced positive emotions while riding a horse, the associative memory has stored representations that link exteroceptive representations (e.g., a visual representation of a horse) and positive interoceptive representations (i.e., positive internal-state patterns). The activation of these representations can induce an emotional reaction in two ways. First, it can lead to an emotional expression through the involvement of the emotion-instructions-generation subsystem and the emotion-execution subsystem that would elicit the actual internal state associated with the previous experience of riding a horse (such an internal state is then processed as described in the next paragraph). Second, it can directly activate the internal-state patterns associated with the previous experience of riding a horse. Once that internal-state pattern is activated, it can participate in making the decision of "riding instead of walking." On the other hand, if negative internal-state patterns were associated with riding a horse, the exteroceptive-interoceptive representations would participate in making the decision of "keeping on walking."

The third process, unlike the ones described above, necessarily leads to a conscious identification of one's own emotional state. Suppose that the individual, while riding a horse, faces a potentially dangerous dog staring at him. The processing of such a situation leads to a diffuse activation of representations in associative memory caused by inputs from many subsystems (not described here; see Kosslyn & Koenig, 1995), reflecting the fact that many aspects of the situation are identified. This can lead to the activation of representations stored in one of the two emotion-instructions-generation subsystems. For example, the behavior of fleeing from the dog involves the generation of withdrawal-related emotion-instructions. Once the instructions are produced, they are converted by the emotion-execution subsystem into an emotional expression (e.g., autonomic, endocrine, facial and postural modifications correlated with the flight behaviour). This emotional expression is an internal state that can be analyzed, like any other perceptive information, by a preprocessing subsystem and a pattern activation subsystem. Once an internal-state pattern is activated (i.e., a particular internal state is recognized as familiar), information sent by the internal-state pattern subsystem to associative memory is completed by information provided by the somatotopic mapping subsystem. If in the context of the representations activated by the situation, the internal state is identified in associative memory, the individual feels the conscious emotion of "being afraid of the dog," because many aspects of the situation, including the emotion, are identified.

Conclusion

A major aim of this paper was to propose a set of principles upon which a computational model of emotion should ideally rely. In this context, we aimed at revisiting the computational architecture of emotion proposed by Kosslyn and Koenig (1995). We can conclude that the model seems fully compatible with both our set of principles and recent data in the domain of emotion. Our approach has led us to develop the model further by specifying some aspects of its architecture. In particular, we reported evidence for distinct internal-state pattern activation subsystems, one storing positive patterns, the other negative ones. Moreover, we reported evidence for distinct emotion-instructions-generation subsystems, one involved in approach behavior, the other in withdrawal behavior.

In conclusion, we would like to emphasize the necessity of mutual fertilization between CN and AI in the making of computational models of emotions. We hope to have shown that the mutual fertilization *within* CN can be productive. We also feel that mutual fertilization *between* CN and AI seems promising in the study of emotions (e.g., Picard, 1997, see chapter 2). Moreover, interactions between Computational Social Sciences (see Castelfranchi, 2001) and Social Cognitive Neuroscience (see Ochsner & Schacter, 2000) might also lead to new insights in modeling social cognition.

In our view, advances in cognitive neuroscience may potentially reduce the gap between research on natural cognition and artificial cognition. Indeed, conducting more advanced studies on natural cognition and producing more complex models is likely to result in more *explicit* and *operational* architectures in order to facilitate simulations.

In this attempt to characterize the cognitive system, an extremely challenging goal is to elaborate architectures of conscious experience, in particular of emotional *experience*. Indeed, in the domain of emotion, cognitive science has just started studying the missing link between the "processing mind-brain" and the "experiencing mind-brain." We are confident of the fact that paradigmatic advances will allow a computational investigation of this link that is essential for an integrated view of emotion.

References

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11,* 231–239.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A.J., Phelps, E. A., Anderson, A., Lee, G. P., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia, 37,* 1111-1117.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience, 20*(7), 2683–2690.

Amaral, D. G., Price, J. L., Pitkänen, A., & Carmichael, S. T. (1992). Anatomical organization of the primate amygdaloid complex. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspect of emotion, memory and mental dysfunction* (pp. 1-66). New York: Wiley-Liss.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411,* 305-309.

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

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269,* 1115-1118.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275,* 1293-1295.

Blair, R. J. R., Morris, J. S., Frith, C. D, Perret, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain, 122,* 883-893.

Borod, J. C., & Madigan, N. K. (2000). Neuropsychology of emotion and emotional disorders: an overview. In J. C. Borod (Ed.), *The neuropsychology of emotion,* (pp. 3-30). New York: Oxford University Press.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E., & Rosen B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*(5), 875-887.

Buchel, C., & Dolan R. J. (2000). Classical fear conditioning in functional neuroimaging. *Current Opinion in Neurobiology, 10*(2), 219-223.

Bush, G., Luu, P., & Posner M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cognitive Science, 4*(6), 215-222.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology, 50,* 191-214.

Cahill, L. (2000). Modulation of long-term memory in humans by emotional arousal: adrenergic activation and the amygdala. In J. P. Aggleton (Ed.), *The amygdala,* (pp. 425-446). New York: Oxford University Press.

Cañamero, L. (2001). Emotions and adaptation in autonomous agents: a design perspective. *Cybernetics and Systems: An International Journal, 32*, 507-529.

Canli, T. (1999). Hemispheric asymmetry in the experience of emotion: A perspective from functional imaging. *Neuroscientist, 5*(4), 201-207.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *The Journal of Neuroscience, 20*, RC99 (1–5).

Castelfranchi, C. (2001). The theory of social functions: challenges for computational social science and multi-agent learning. *Journal of Cognitive Systems Research, 2*, 5–38.

Clore, G. L. (1994). Why emotions are felt. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: fundamental questions*, (pp. 103-111). New York: Oxford University Press.

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2001). Neuroanatomical basis for first and second-order representations of bodily states. *Nature Neuroscience, 4*(2), 207-212.

Crosson, B., Radonovich, K., Sadek, J. R., Gökçay, D., Bauer, R. M., Fishler, I. S., Cato, M. A., Maron, L., Auerbach, E. J., Browd, S. R., & Briggs, R. W. (1999). Left hemisphere processing of emotional connotation during word generation. *ReuroReport, 10*, 2449-2455.

Damasio, A. R. (1994). *Descartes' error*. New York: Grosset/Putnam Book.

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Research Reviews 26*, 83–86.

Davidson, R. J., (1992). Emotion and affective style: Hemispheric substrates. *Psychological Science, 3*, 39-43.

Davidson, R. J. (1995). Cerebral asymmetry, emotion, and affective style. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry*, (pp. 361-387). Massachusetts: MIT Press.

Davidson, R. J. (2000). The functional Neuroanatomy of affective style. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion*, (pp. 371-388). New York: Oxford University Press.

Davidson, R. J., & Irwin. W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences, 3*(1), 11-21.

Davidson, R. J., Jackson, D. C., & Kalin, N. H. (2000). Emotion, plasticity, context, and regulation: perspectives from affective neuroscience. *Psychological Bulletin, 126*(6), 890-909.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: Murray

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of the striate cortex. *Neuroreport, 10*, 3759-3763.

Descartes, R. (1649). *Les passions de l'âme*. Paris.

Devinsky, O., Morrell, M. J., & Vogt B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain, 118*(1), 279-306.

Dolan, R. J. , Morris J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences, 98*(17), 10006-10010.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.). *Handbook of cognition and emotion*. London: John Wiley & Sons, Ltd.

Ekman, P., Levenson, R. W., & Friesen, W. W. (1983). ANS activity distinguishes among emotions. *Science, 221*, 1208–1210.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitude. *Journal of Personality and Social Psychology, 50*, 229-238.

Frijda, N. H. (1994). Emotions are functional, most of the time. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions,* (pp. 112-122). New York: Oxford University Press.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates.

Gainotti, G. (2000). Neuropsychological theories of emotion. In J. C. Borod (Ed.), *The Neuropsychology of Emotion*, (pp. 214-238). New York: Oxford University Press.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences*, *2*(12), 493-501.

Gazzaniga, M. S. (2000). *The new cognitive neurosciences*. Massachusetts: MIT Press.

George, M. S., Ketter, T. A., Parekh, P. I., Herscovitch, P., & Post, R. M. (1996). Gender differences in regional cerebral blood flow during transient self-induced sadness or happiness. *Biological Psychiatry, 40*, 859-871

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5*(9), 394-400.

Hamann, S., & Hui, M. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *NeuroReport, 13*(1), 15-19.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *Neuroreport, 11*(1), 43-48.

Hermans, D., de Houwer, J., & Eelen, P. (1994). The affective priming effect : Automatic activation of evaluative information in memory. *Cognition and Emotion, 8*(6), 515-533.

Hobbes, T. (1651/1985). *Leviathan*. England: Penguin Classics.

Irwin, W., Davidson, R. J., Lowe, M. J., Mock, B. J., Sorenson, J. A., & Turski, P. A. (1996). Human amygdala activation detected with echo-planar functional magnetic resonance imaging. *NeuroReport, 7*, 1765-1769.

James, W. (1884). What is an Emotion? *Mind, 9*, 188-205.

Ketter, T. A., Andreason, P. J., George, M. S., Lee, C., Gill, D. S., Parekh, P.I., Willis, M. W., Herscovitch, P., & Post, R. M. (1996). Anterior paralimbic mediation of procaine-induced emotional and psychosensory experiences. *Archive of General Psychiatry, 53*, 59-69.

Klauer, K. C., & Musch, J. (2002). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion*. Mahwah. NJ: Lawrence Erlbaum.

Kleinginna, P. R. Jr., & Kleinginna, A. M. (1981). A categorized list of emotion definitions with suggestions for a consensual definition. *Motivation and emotion, 5*, 345-379.

Kosslyn, S. M., & Koenig, O. (1995). *Wet mind, the new cognitive neuroscience*. New York: Free Press.

Kosslyn, S.M., Shin, L.M., Thompson, W.L., McNally, R.J., Rauch, S.L., Pitman, R.K., & Albert, N. M. (1996). Neural effects of visualizing and perceiving aversive stimuli: a PET investigation. *Neuroreport, 7*,1569-1576.

LeDoux, J. E. (1994). Emotional processing, but not emotions, can occur unconsciously. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions,* (pp. 291-293). New York: Oxford University Press.

LeDoux, J. E. (1996). *The emotional brain*. London: Weidenfeld & Nicolson.

Maddock, R. J., & Buonocore, M. H. (1997). Activation of left posterior cingulate gyrus by the auditory presentation ofthreat-related words: an fMRI study. *Psychiatry Research: Neuroimaging Section 75*, 1-14.

Maratos, E. J., Dolan, R. J., Morris, J. S., Henson, R. N. A., & Rugg M. D. (2001). Neural activity associated with episodic memory for emotional context. *Neuropsychologia, 39*, 910–920.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, L. McGaugh & N. M. Weinberger (Eds.), *Neurobiology of Learning and Memory*. New York: Guilford Press.

Morris, J. S., Frith, C. D., Perret, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdalia to fearful and happy facial expressions. *Nature, 383*, 812-815.

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the amygdala. *Nature, 393*, 467-470.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl. Acad. Sci. USA, 96*, 1680-1685.

Morris, J. S., Buchel, C., & Dolan, R. J. (2001). Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. *Neuroimage, 13*(6 Pt 1), 1044-52.

Morris, J. S., de Gelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain, 124*(6), 1241-52.

Ochsner, K. N., & Schacter, D. L. (2000). A social cognitive neuroscience approach to emotion and memory. In J. C. Borod (Ed.), *The neuropsychology of emotion*, (pp. 163-193). New York: Oxford University Press.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522.

Öhman, A., & Soares, J. J. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. *Journal of Experimental Psychology General, 127*(1), 69-82.

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia, 36*(8), 701-15.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*(9), 337-344.

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001) Activation of the left amygdala to a cognitive representation of fear. *Nature Neuroscience., 4*(4), 437-41.

Phillips, A. G. (1999). "The brain and emotion" by Edmund T. Rolls. *Trends in Cognitive Sciences, 3*(7), 281-282.

Picard, R. (1997). *Affective computing*. Massachusetts: MIT press.

Popper, K. R. (1968). *The logic of scientific discovery* (Revised edition). Hutchinson: London.

Posner, M. I., & DiGirolamo, G. J. (2000). Cognitive neuroscience: Origins and promise. *Psychological Bulletin, 126*(6), 873-889.

Rolls, E. (1999). *The brain and emotion*. New-York: Oxford University Press.

Sander, D., & Koenig, O. (submitted). *Hemispheric asymmetry in explicit and implicit emotional evaluation of pictures: A divided visual field study.*