Reflections on Emotion

Jan Morén and **Christian Balkenius**

Lund University Cognitive Science

Kungshuset, Lundagrd

S-222 22 Lund

Sweden

email: jan.moren@lucs.lu.se

christian.balkenius@lucs.lu.se

Abstract

When trying to model autonomous agents, the emotional aspects of motivation are often overlooked. We discuss the role emotions play in this field and present a framework that ties together emotions, motivations and action selection.

1 Introduction

Autonomy is often defined as the ability of an agent to operate without supervision. This, however, is too broad since it will include a lot of systems that react predictably driven by externally imposed goals such as industrial robots and toasters. Nor is it sufficient that the agent can make decisions on its own or that its behavior is based on its own experience. Instead, we propose that autonomy is the ability to generate goals internally, based on the needs of the system. These needs may, of course, be the result of design which makes it possible to control an autonomous system.

At first sight, the difference between defining the needs of a system and defining its goals may seem slight, even nonexistent. However, we argue that there is a major difference between these two approaches. Whereas goals (as we speak of them) are task-specific, needs are a minimal set of objectives needed for the system to successfully exist in its environment. It is the difference between "wanting to find food" and "being hungry".

Defining goals for a system implies setting a predetermined prioritation among the possible activities the system is expected to perform. Any flexibility regarding the apropriateness for pursuing a given goal must be explicitly or implicitly built in by the designer. The other problem associated with explicit goals is that although the system may be able to generate subgoals, it cannot generate entirely new goals if the situation demands it. In short, setting goals means sacrificing flexibility and adaptability for control; the system is not fully autonomous.

Giving the system needs, then allowing the systems to generate internal goals, on the other hand, means that we give the system maximal opportunity to fulfill its mission in any way it sees fit. As we describe below, the needs are evaluated together with an emotional evaluation to generate an objective, which will subsequently drive action selection. Of course, as needs are rather more abstract, the designer would sacrifice some control over the system to achieve a greater degree of autonomy.

There has traditionally been a sharp divide between the concepts of learning and reasoning on one hand and the concept of emotions on the other. This is in no small part due to the relative ease with which learning can be studied and with the more obvious usefulness of understanding the mechanisms of learning.

The traditional dichotomy between cognition and emotion is probably responsible for the lack of interest in emotional and motivational theories within cognitive science [LeDoux, 1995]. This is unfortunate for a number of reasons [Balkenius, 1995]. First, there seems to be no clear distinction between cognition, motivation and emotion in our daily activities. It appears that all learning is biased by emotive values. Next, learning cannot be understood without reference to motivation. Higher cognitive processes like categorization and problem solving only play any role when they are used to fulfil some goal of the individual, that is, only when they are motivated. This is unfortunately an aspect of learning that is often ingnored. Last, we argue that the role of emotions is to tell the animal what it should have done, when its expectations are inaccurate or its behavior is unsuccesful.

2 Fundamentals

One major problem with dealing with emotional processes is the confusing terminology associated with this area. Terms properly associated with emotions have been expropriated for use in other fields and, not uncommonly, the same phenomenon have been given different labels depending on the field in question.

When we talk about emotions in this paper, we do not discuss the subjective feeling that we experience, but an *evaluation* of a stimulus as being emotionally charged. Such emotions are either *primary* or *secondary*.

Primary emotions are generated by stimuli or contexts directly and intrinsically related to the needs of the system. These can be things like the smell of food, pain or sexual signals. These stimuli do not need to

be associated with other emotional stimuli and are resistant to any change in their effectiveness. However, their expression can be inhibited by other systems (see Section 8).

Secondary (or higher order) emotions are stimuli that are not emotionally charged in themselves, but that becomes emotional through association with a primary emotional stimuli (where external reinforcers would count as such), or with secondary stimuli already emotionally charged. This association enables these stimuli to act the same as primary stimuli, being used for motivation, as well as directing attention. In effect, secondary emotions predict the occurrence of primary emotions.

Although the emotional system will react to both pleasant and unpleasant stimuli, most of the work in this area has been focused on fear [LeDoux and Fellous, 1995]. Fear can be defined as the reaction to a signal that predicts punishment. This signal is said to be *aversive*. Fear in this sense is often equated with anxiety [Gray, 1982]. Fear as an emotional state gives rise to *avoidance* behaviors. When an animal feels fear it will try to avoid whatever made it fearful.

When a system does not receive an expected reinforcer (or positively charged stimulus), the result is *frustration*, and anger is a reaction to this [Rolls, 1995]. If the system bumps into a wall unexpectedly, for instance, this may make it frustrated since it was not able to move closer to its goal which could result in agressive behavior towards the wall or anything else that happens to be around at the moment.

Positive emotions are less well defined but can be called hope or anticipation [Panksepp, 1981]. These emotions are reactions to *appetitive* stimuli that are rewarding or predict reward [Rolls, 1995].

Stimuli can also become emotionally charged in the abscence of primary stimuli, if they are unexpected [Gray, 1982]. They will arouse interest and direct attention to them for further evaluation. The emotional reaction to novelty is both aversive and appetitive; this is an approach-avoidance conflict caused by a lack of experience with the stimulus [Lewin, 1936].

Common terms associated with emotions are *motivation* and *drive*. Although sometimes thought of as the same thing, they are in fact rather distinct. Hebb [1955] popularized the drive concept and described a model accounting for its function. He sees a drive as motivational energy driving the organism towards activity while leaving the nature of the needed activity largely unspecified. He also relates the concept of drive to reward and punishment. An optimal drive level is rewarding while a too high or low drive is punishing. Since novelty increases drive, a moderate level of novelty is rewarding.

Motivation is in our view much more focused than drive; it is the combination of internal needs, emotional state and external context [Balkenius, 1995]. It does not only describe what to accomplish, but also in part how to do it.

So, what about the feelings we associate with emotions? A common theory, described by [LeDoux, 1996] among others, states that the subjective feeling we experience is generated as a result of our reacting to a pleasant or unpleasant stimuli. There is some evidence that it in fact is the *reaction* itself that elicits this subjective sensation [Schachter, 1969].

3   A Model of Emotional Integration

We will not discuss details of the emotional system itself in this paper (although there is a section on biological relevance below); for such a discussion, see [Balkenius and Morén, 1996; LeDoux and Fellous, 1995; Rolls, 1986]. Instead we give a system-level description of the proposed architecture for an autonomous, emotionally driven system.

In figure 1 we have a principal view of the emotional subsystem as a part of a larger autonomous system. We have here omitted the subsystems dealing with the emotional aspects of attention and long term memory acquisition, concentrating instead on action selection.

This description focuses on three interacting subsystems. The *motivational system* compares the internally generated needs with the emotional evaluations and generates *objectives* for action selection.

The *action selection system* uses the objectives to generate action sequences, changing the internal state of the system in the process. This system also receives information about outside stimuli and can create complex actions. If the objective is too abstract, the changes in internal context will enable the emotional and motivational subsystems to create more concrete objectives to accomplish the main objective.

Using the external stimuli and the external and internal contexts, the *emotional system* evaluates the stimuli for the motivational subsystem [Balkenius, 1995]. This subsystem can also generate emotional reactions directly without involving the action selection system.

This model is a form of two-process model, as proposed by Rolls [1986] and Mowrer [1960]. One process, the emotional subsystem, learns an evaluation of stimuli thus forming an opinion of the desirability or undesirability of the stimulus. The emotional subsystem subsequently generates an objective (or general response) to deal with this stimulus, which the action subsystem carries out as best it can.

4   Motivation

The first of the three components is the motivational subsystem. This system receives the internal needs and the emotional evaluation of the present stimuli and context. This evaluation influences the relative importance of the present needs and allows for the motivational system to generate an objective for the action system. The needs are internally generated, and correspond very roughly to Hebb's drives [Hebb, 1995].

The comparison of the internal needs and the emotional evaluation of stimuli enables the motivational subsystem to take into account the current situation when choosing an objective; a system that went into the kitchen because it was hungry might well take a sip of water as well when close to the faucet, even
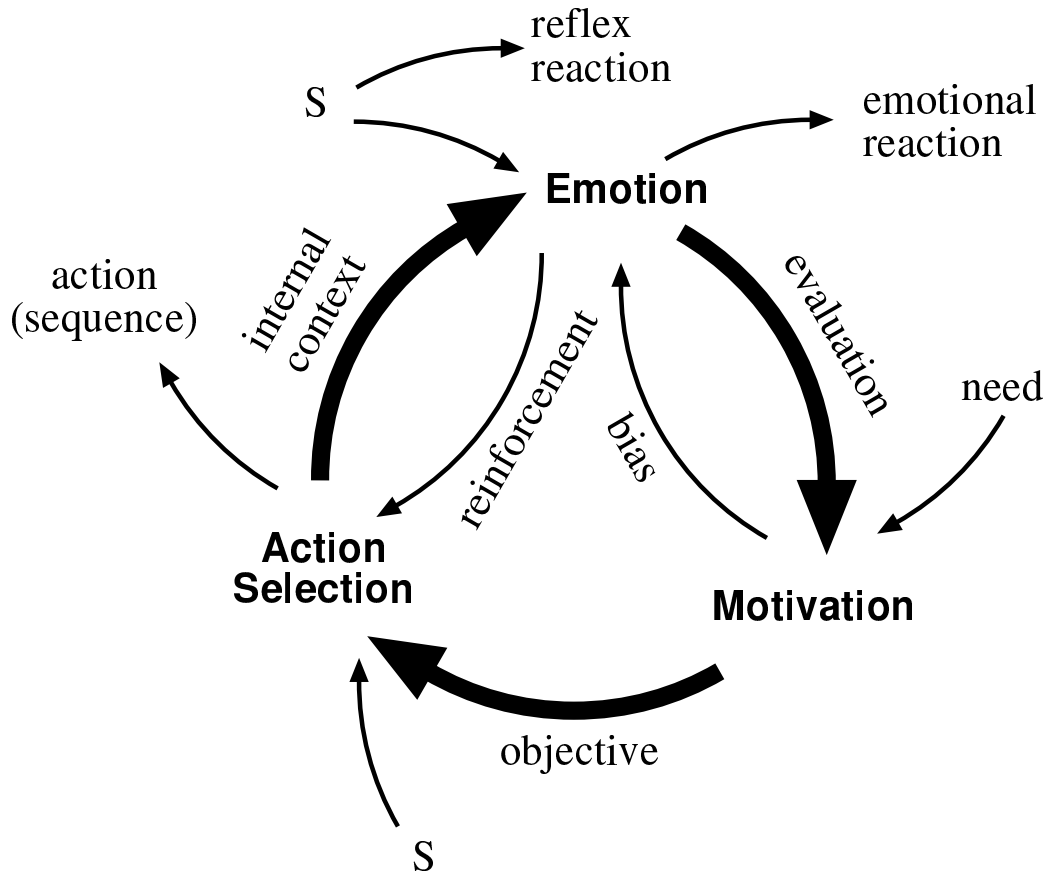
Figure 1: A model for generating actions based on internal needs and emotional evaluation of stimuli. This model has three interconnected main subsystems, incoming data from sensory subsystems and internal context, and outgoing paths leading to other subsystems (not described here).

though it was nowhere near as thirsty as it was hungry. This is an example of oppurtunistic behavior [Balkenius, 1993]. On the other hand, when a need is completely fulfilled, the system would normally not react to a positive emotional evaluation unless it was very strong. When satiated, a system would not eat a sandwich lying in front of it, but might eat a small piece of candy.

We use the term 'objective', rather than 'goal'. This is mainly because 'goal' connotates a concrete, closed outcome, rather than a more abstract desire or state. An objective *can* be very concrete ("eat that sandwich") and will generate concrete actions, but can also be a high-level desire that in turn will determine one or several other objectives that eventually will fulfill this one.

The motivational system also outputs a bias signal back to the emotional system which is used in the evaluation. For example, food is only important if you are hungry [Rolls, 1995].

5   Action

Action selection uses the motivational objective, the present stimuli and the context to generate actions to resolve the objective. These actions can be highly structured and context dependent; this subsystem is

able to do a great deal of planning within the present context. The outcome of this system can be twofold. First, this system generates an action sequence if it is able to; these actions will of course in turn change the present stimuli and the externally generated context. If, however, the objective is too abstract or dependent on long-term memory, no explicit actions will be generated.

The system will also generate an internal context that in turn will influence both the emotional system and the action system itself. This internal context consists of needs, short term – or active – memory (including cognitive structures), bodily states and emotional state.

The action system will also get an evaluation directly from the emotional system. This evaluation is used as reinforcement when the action system learns to perform motor sequences.

6   Emotions

The emotional subsystem will make use of the stimuli and both external and internal contexts to evaluate stimuli. This evaluation is in one of two forms: first order and second order emotional evaluation.

First order – or primary – evaluation occurs with stimuli that are intrinsically emotionally charged

(such as pain), but unexpected stimuli are also charged, as they can have a potentially large impact on the system. Many of these stimuli will by themselves generate an emotional reactions directly independently of action selection. Primary stimuli can also generate reflex reactions such as withdrawal even before they enter the emotional system.

Secondary stimuli are those that by themselves do not elicit a reaction, but acquires emotional content through association with a primary or another secondary stimulus. There is ample evidence that the learning being performed here can be described as classical conditioning [LeDoux, 1992].

The emotional system uses the current context to be able to rightly evaluate emotional stimuli, and the systems' needs are certainly a part of the internally generated context.

The emotionally charged stimuli are used in several ways. First, as we have described here, the evaluation is used as *incentive motivation* by the motivational system to produce objectives. It is also used as a reinforcer for motor-learning in the action-selection system. Additionally, they are used for long term memory and attention, as we will see in the next section.

7   Learning, Memory and Attention

Mowrer [1960] established a two-process model whereby an emotional system evaluated stimuli and the evaluation then being used in the learning system proper. By not only advocating this role of emotions in learning, but also suggesting how such a system could be implemented, this work spawned a good deal of interest and the development of several new models based on this idea (see [Gray, 1975; Klopf, 1988]).

One of the primary functions of emotions is the capability of evaluating stimuli. When a previously unknown or unremarked stimuli occurs in association with an emotionally charged stimuli, the emotional system will associate this new stimuli with the same or a similar emotional content. Traditional learning methods all rely on some form of reinforcer, presumably generated from the outside. In reinforcement learning methods that allow for internally generated reinforcement, it is still very directly linked to the external reinforcer and the problem has been reduced to one of credit assignment [Kaelbling *et al.*, 1996], as the reinforcers bacome directly linked with the specific actions taken by the system at the time.

In real life, of course, the situation is much more complicated; specifically, solving this as a credit assignment problem will not enable the system to transfer hard-won knowledge between contexts. Once a stimuli is evaluated by the emotional system, this evaluation can then be used as a basis both for evaluation of other stimuli and for evaluation of the contexts themselves.

The second function of the emotional system is to focus the system's attention where it would do the most good. The world is too complicated, and the sensory subsystems too variegated, for the system to be able to spend time and other resources on it all. By using the emotional systems' capability for evaluation and prioritation, this sensory barrage can be sifted through so only the most relevant stimuli recieves any attention. Very closely related to this is of course the need to decide what stimuli to retain as long term memories, and we believe this is accomplished by the same mechanism.

8   Biological Foundations

We will here only have a brief discussion of the neurological foundations underlying emotional encoding; for an extended view, see cf. [Aggleton, 1992].

The amygdala, a small almond-shaped region in the temparal lobe, has been identified as crucial for emotional reactions, motivation, learning and memory [LeDoux, 1992; Rolls, 1986]. As such, it forms an important part of the emotional subsystem.

Stimuli can get emotionally charged in the abscence of primary stimuli as well, if they are unexpected; they will thus arouse interest and direct attention to them for further evaluation.

The assumption here is that the amygdala reacts to stimuli that are either emotionally charged (causing fear or discomfort) or to stimuli that are novel, i.e. that have not been encountered recently and that have not previously been associated with a charged stimuli. When such a stimuli occurs in the right temporal proximity to a charged stimuli or to a primary reward, this association will be learned.

Working closely with the amygdala is the orbitofrontal cortex (OFC), which will evaluate the activity of the amygdala in context. In the event of an omitted reward or punishment, this area will inhibit areas of the amygdala, as well as other areas [Rolls, 1995], thus inhibiting the associations responsible for the expectation of a reward/punishment. Together, the amygdala and orbitofrontal areas make up the emotional system of our description above.

Among other areas, the amygdala projects into various parts of hypothalamus. This area is thought to be heavily involved with innate needs such as hunger [schachter, 1970]. Hypothalamus also projects back into the amygdala. We suggest that this region corresponds to our motivational system.

The basal ganglia and the motor cortex are thought to be responsible for learning and executing motor sequences. The basal ganglia receives projections from amygdala, the sensory cortex and the frontal cortex, giving primaray reinforcement, sensory information and context, respectively. This area projects to the motor cortex and back into thalamus and the prefrontal cortex. The basal ganglia and motor cortex is the aforementioned actions selection system.

Finally, the hippocampus is thought to generate and store contextual information needed by other areas [Nadel *et al.*, 1985). It has extensive interconnections with the sensory cortical areas and also projects to the orbitofrontal cortex.

9 Conclusion

The emotional system is an important part of many functions, among them learning, attention and long term memory; here we have concentrated on the role of emotions in motivation and action selection.

We have discussed a framework for emotionally based action. The framework consists of three inter-related subsystems that are tightly coupled. The motivational subsystem creates objectives for the system based on internal needs and the emotional evaluation of current stimuli; these objectives are carried out by the action selection system.

In conclusion, the motivational system informs the system what should be done, the action system reports what was done, and the emotional system indicates what should have been done.

Acknowledgements

References

[Aggleton, 1992] John P. Aggleton, editor. *The Amygdala: neurobiological aspects of emotion, memory, and mental dysfunction.* Wiley-Liss, New York, 1992.

[Balkenius, 1993] Christian Balkenius. Motivation and attention in an autonomous agent. Presented at *the workshop on architectures underlying motivation and emotion - WAUME 93.* University of Birmingham, Birmingham, 1993.

[Balkenius, 1995] Christian Balkenius. Natural intelligence in artificial creatures. Lund University Cognitive Studies 37, Lund, 1995.

[Balkenius and Morén, 1998] Christian Balkenius and Jan Morén. (1998). A Computational Model of Emotional Conditioning in the Brain. In Dolores Canamero, Chisato Numaoka and Paolo Petta, editors, *Workshop on Grounding Emotions in Adaptive Systems.* Zürich, 1998.

[Gray, 1975] John A. Gray. *Elements of a two-process theory of learning.* Academic Press, London, 1975.

[Gray, 1982] John A. Gray. *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system.* Oxford University Press, Oxford, 1982.

[Hebb, 1955] Donald O. Hebb. Drive and the C. N. S. *Psychological Review,* 62(4): 243–254, 1955.

[Houk *et al.*, 1995] James C. Houk, Joel Davis and David Beiser, editors. (1995). *Models of Information Processing in the Basal Ganglia.* MIT Press, Cambridge, MA, 1995.

[Kaebling *et al.*, 1996] Leslie P. Kaelbling, Michael L. Littman and Andrew W. Moore. Reinforcement Learning; A Survey. *Journal of Artificial Intelligence Research,* 4: 237–285, 1996.

[Klopf, 1988] A. Harry Klopf. A neuronal model of classical conditioning. *Psychobiology,* 16(2): 85–125, 1988.

[LeDoux, 1992] Joseph E. LeDoux. Emotion and the Amygdala. in John P. Aggleton, editor, *The Amygdala:neurobiological aspects of emotion, memory, and mental dysfunction* , pages 339–351, New York, 1992. Wiley-Liss.

[LeDoux, 1995] Joseph E. LeDoux. In search of an emotional system in the brain: leaping from fear to emotion and consciousness. In Michael S. Gazzaniga, editor, *The cognitive neurosciences,* pages 1049–1061, Cambridge, MA, 1995. MIT Press.

[LeDoux and Fellous, 1995] Joseph E. LeDoux and Jean-Mark Fellous. Emotion and Computational Neuroscience. In Michael A. Arbib, editor, *The handbook of brain theory and neural networks,* Cambridge, Mass., 1995. MIT Press.

[LeDoux, 1996] Joseph E. LeDoux. *The Emotional Brain.* Simon & Schuster, New York, 1996.

[Lewin, 1936] Kurt Lewin. *Principles of topological psychology.* McGraw-Hill, New York, 1936.

[Mowrer, 1960] Orval H. Mowrer. *Learning theory and behavior.* Wiley, New York, 1960.

[Nadel, 1985] Lynn Nadel, J. Willner and E. M. Kurz, E. M. Cognitive maps and environmental context. In Peter D. Balsam and A. Tomie, editors, *Context and learning,* pages 385–406, Hillsdale, NJ, 1985. Erlbaum.

[Panksepp, 1981] Jaak Panksepp. Hypothalamic integration of behavior. In Peter J.Morgane and Jaak Panksepp, editors, *Handbook of the hypothalamus,* New York, 1981. Marcel Dekker.

[Rolls, 1986] Edmund T. Rolls. A theory of emotion, and its application to understanding the neural basis of emotion. In Y. Oomura, editor, *Emotions: neural and chemical control,* pages 325–344, Tokyo, 1986. Japan Scientific Societies Press.

[Rolls, 1995] Edmund T. Rolls. A theory of emotion and consciousness, and its application to understanding the neural basis of emotion. In Michael S. Gazzaniga, editor, *The cognitive neurosciences,* pages 1091–1106, Cambridge, MA, 1995. MIT Press.

[Schachter, 1964] Stanley Schachter. The interaction of cognitive and physiological determinants of emotional state. In Leonard Berkowitz, editor, *Advances in experimental social psychology,* pages 8–49, New York, 1964. Academic Press.

[Schachter, 1970] Stanley Schachter. Some extraordinary facts about obese humans and rats. *American Psychologist,* 26:129–144, 1970.