

Beyond Needs: Emotions and the Commitment Requirement

Michel Aubé
Université de Sherbrooke
Michel.Aube@USherbrooke.ca

In modelling the mind, Aaron Sloman (1995) used to say, «architecture is far more important than design». Moreover, robust computational architecture has to stem from an adaptive mapping from «requirement space» onto «design space». Indeed, for any computational system to «survive» or succeed in a given environment, there is a set of requirements to be met, and to meet them, there is a whole space of possible designs which could satisfy the constraints they exert on survival or success. Mind modelling thus implies that a reasonable set of requirements be clearly specified for a given «species» (or class of systems), and that a variety of possible designs be explored as reasonable solutions for handling those constraints.

The following chapter is to argue that emotional systems probably emerged through natural selection as a well delineated class of designs, slowly carved, patch by patch, to handle very peculiar kinds of requirements. It is the author's opinion that in most current computational models of the mind, emotions are unfortunately merged with a whole bunch of different motivational processes of differing kinds, functions and levels of complexity, and that this confusion has considerably obscured their successful understanding and implementation. Yet, such an assertion does not presuppose that emotions have all emerged at about the same evolutionary period, nor even in close succession. Just as the different perceptual systems probably emerged more or less randomly, provided they offered adaptations significant enough to be selected, so the different emotional systems probably also issued from random mutations that proved successful for concerned individuals, given the requirements they had to face in their own ecosystem. Yet, it is quite profitable for an engineer to conceive of «any» perceptual system as an «energy transducer» which could thereby provide information for a given artefact. Such a theoretical view certainly helps to design and implement one such system. In a similar way, we contend that emotional systems could all be envisioned as belonging to a class of designs built around certain shared principles to fulfil a certain class of functions. But what are those specific requirements that any kinds of emotions are presumably designed to handle?

There are many independent yet converging indications that most emotional episodes are plainly «interactive» in character. One of these indications certainly has to do with expressions. All emotions seem indeed to incorporate an expressive aspect, out of which the feeling itself does not seem quite complete (Ekman, 1982, 1984; Evans, 2002; Frijda, 1989). It is clearly the case for laughing (Provine, 2000), crying (Lutz, 2001), empathic distress (Levenson and Ruef, 1992; Sagi and Hoffman, 1976), shame and guilt (Baumeister *et al.*, 1994; Lewis, 1993; Tangney, 1995), fear (Hoffman, 1974; Rosenblum and Alpert, 1974), anger (Averill, 1979, 1983; Lerner and Dodge, 1993)... Expressions are hard to control, as if they were an intrinsic part of emotion, and when they are successfully retained, the feeling itself often almost dies away (Ekman and Davidson, 1994). Their control remains one of the most successful path to the socialisation of emotions and to the mastering of display rules (Ekman, 1993; Malatesta and Haviland, 1982). On

the other hand, actors well know that putting up the face or posture of a given emotion is a powerful way of eliciting it (Bloch, 1989; Bloch *et al.*, 1987). Finally, organisms that have emotions are all subjects to strong contagion from the expression of emotions in others (Dimberg, 1982; Hatfield *et al.*, 1994; Klinnert *et al.*, 1983). Now, all this suggests that emotions might have emerged basically as communicative devices (Oatley and Johnson-Laird, 1996), presumably designed to resolve certain survival problems precisely through such acts of communication.

Another intriguing indication is that most emotion theorists will recognise that emotions are more frequently triggered in situations of encounter, that they are also most intense when interpersonal relationships are involved, and yet all the more when these relations are intimate. Herbert Simon (1967) already mentioned this in his seminal paper on «Motivational and Emotional Controls of Cognition», and many psychological experiments and surveys have amply confirmed this ever since (Boucher, 1983; Scherer, 1988; Scherer *et al.*, 1986). Finally, some neuroscientists following Paul MacLean (1993; Damasio, 1994, 1999, 2003; LeDoux, 1998; Panksepp, 1991) contend that emotions largely depend upon limbic structures, but also that maternal behaviour, distress cries, imprinting and attachment all appeared in evolution conjointly with those neural structures. Many psychologists and ethologists following Bowlby (1969, 1973, 1980; Ainsworth, 1989) and Eibl-Eibesfeldt (1975) will further suggest that parent-offspring attachment is the paradigm scenario of all successive collaborative behaviours. All of these considerations lead us to envision the «need to belong» (Baumeister and Leary, 1995) as a necessary element for the understanding of emotions and probably even as a core component of what they are, what they are for (Aubé and Senteni, 1996b) and how they manage to play their part.

How do we put all this together? One way is first to distinguish emotions as a subclass of motivations. We thus envision the motivation construct as the «psychic force» that lay behind and give sense to all changes in behaviour (Vallerand and Thill, 1993). Motivational systems play against psychic inertia. They have evolved to make the organisms change from behaviour to behaviour when presumably more adaptive to do so. One key to their operation has to do with managing «resources» critical for survival. For instance, a need system such as hunger is built so as to detect deficits in nutriments and to set the whole organism to look for the appropriate resource. Need systems are understood well enough in Biology and Psychology (Toates, 1986). They all consist in feedback systems that monitor certain specific resources and trigger certain behaviours whenever it appears required to do so. As a first approximation, we propose that all kinds of motivations rest on a similar control loop structure. Now, if emotions appear as some peculiar kinds of motivations, a fundamental question pops out as to what kind of resources they manage. What has been said above concerning the interactive character of emotions points to the idea that it might have to do with something like bonding, belonging, and collaborative behaviour.

Now, there is one subtlety here. One might think that the involved resource has to do with profiting from other agents. But non collaborative others are of no use here. What indeed appears critical has to do with the «predisposition to collaborate in a reciprocal manner». We call «commitment» this predisposition to collaborate and be helpful, provided the other will reciprocate in turn (Frank, 1988; Trivers, 1971). This concept comes from a long tradition in sociology and distributed AI, and it seems to offer a solid ground for conceptualising emotions (Becker, 1960; Dongha, 1994; Fikes, 1982; Gasser, 1991; Gerson, 1976; Gouldner, 1960; Jennings, 1993; Kerr and Kaufman-Gilliland, 1994). Yet we put this concept to a special use in

our model (Aubé, 1997, 1998, 2001; Aubé and Senteni, 1995, 1996a, 1996b). We call these commitments second-order resources, to distinguish them from simpler ones such as food or water, and we contend that, in nurturing species such as birds or mammals, they are often as critical as first-order resources. For instance, new-borns and offspring in these species certainly would not survive without them. In highly social species such as ours, commitments are just the more essential. Just as needs are computational systems that handle the management of first-order resources, we propose that emotions are computational systems that handle the management of second-order resources (or commitments) that are at the very root of «sociality» (Castelfranchi, 1990). The challenge of mastering these powerful resources likely became a strong selective pressure for species which could profit from collaboration (Krebs, 1987; Nesse, 2001). Such an analysis does not aim at rejecting needs or simpler motivational systems from the model of mind. It rather stresses that a new level of requirements had to be met with the advent of commitments, and that specific kinds of design had to be implemented to handle the more complex behaviours that resulted from their emergence. As was already suggested by Norman (1980), we think that emotions do form kind of an intermediate layer of design in between reactive processes and more reflective ones, and that they probably even contributed to the emergence of the more complex structures which developed on top of them.

We indeed suspect that the emergence of emotions as commitments operators had dramatic consequences in evolution. For one, and not the least, we see in them one of the basic roots for identity. Psychologists such as George Herbert Mead (1934) rightly envisioned that the Self emerged from social interactions, when the individual gets to see himself and his own role from the point of view of the others, and when he incorporates this external stance to his own. Now the management of commitments, which emerged as a prerequisite for attachment behaviour, also requires that individuals be differentiated from each other as unique identifiable partners. It is usually not in the best interest of animal parents that they confuse their own offspring with those of others and invest too much in nurturing them. Nor is it desirable that offspring get attached to their predators! Hence, the very structure of commitments management calls for the emergence of identities. Moreover, if one is to cheerily protect its own commitments to others, it has to reflect upon the consequences of its own behaviours on these commitments. Thus envisioned, emotions also appear at the root of sociality and even of moral behaviour (De Wall, 1996; Tappolet, 2000).

Finally, such an analysis also bear some radical consequences on the implementation of artefacts which are eventually to have «emotions». First of all, emotional systems are definitively more complex than other motivational systems such as needs, in that they answer to more complex requirements. Our analysis thus enables us to distinguish between different kinds of behaviours that are often unfruitfully merged together. Fear is a good example. For instance, we think that what makes a frog jump away into water when your shadow covers it, is much simpler than what makes the frightened kid call for his parents, or what makes the adult give an alarm call to conspecifics in the presence of danger. Actually, we would suggest that the frog's kind of «fear» is «infra-emotional» and rather belongs to the need level. Our model thus postulates that different reactions, which are usually lumped under the same emotional label, very likely belong to different control mechanisms, with different evolutionary histories, and that they are computed by different brain circuits. Even in humans, animal phobias, for instance, are probably much more primitive than social phobias. They react to different kinds of drugs, and Öhman (1986, 1993) has suggested that they stem from an older «predatory-defence system», while social fears rather result from a «dominance-submissiveness system». Such a difference calls for different designs,

and involves the management of quite different resources that could not be detected and handled with the simpler devices. Another consequence of utmost importance is that we think it is useless and ludicrous to try design emotional systems for artefacts that do not belong to communities, within which they have their own well delineated identity. It would be like paying the cost of implementing such a complex behaviour as language within an artefact that would never have to communicate with anything else! Our analysis also commands that commitments themselves be clearly represented within emotional robots as dynamic computational identities (such as «actors», for example - Agha, 1986) that could detect events which could threaten their integrity, or which could sustain and strengthen it. Members of such a community should be built along a similar ontology, so they could not escape from being moved (i.e. motivated) whenever some of their second-order resources are perceived as suddenly threatened or unusually favoured.

References

Agha, G. (1986). *Actors*. Cambridge, MA : MIT Press.

Ainsworth, M. D. S. (1989). Attachments Beyond Infancy. *American Psychologist*, 44, 709-716.

Aubé, M. (1997). Toward Computational Models of Motivation: A Much Needed Foundation for Social Sciences and Education. *Journal of Artificial Intelligence in Education*, 8(1), 43-75.

Aubé, M. (1998). A Commitment Theory of Emotions. In Proceedings of the 1998 AAAI Fall Symposium «*Emotional and Intelligent: The Tangled Knot of Cognition*», Orlando, Florida, 23-25 October 1998.

Aubé, M. (2001). From Toda's Urge Theory to the Commitment Theory of Emotions. *Grounding Emotions in Adaptive Systems*, Special Issue of *Cybernetics and Systems: An International Journal*, 32(6), 585-610.

Aubé, M. and Senteni, A. (1995). A Foundation for Commitments as Resource Management in Multi-Agents Systems. In T. Finin and J. Mayfield, (Eds.), *Proceedings of the CIKM Workshop on Intelligent Information Agents*. Baltimore, Maryland, December 1-2 1995.

Aubé, M. and Senteni, A. (1996a). Emotions as Commitments Operators: A Foundation for Control Structure in Multi-Agents Systems. In W. Van de Velde and J. W. Perram, (Eds.), *Agents Breaking Away*, Proceedings of the 7th European Workshop on MAAMAW, Lecture Notes on Artificial Intelligence, No. 1038, (pp. 13-25). Berlin: Springer.

Aubé, M. and Senteni, A. (1996b). What are Emotions For? Commitments Management and Regulation Within Animals/Animats Encounters. In P. Maes, M. Mataric, J.-A. Meyer, J. Pollack, and S. W. Wilson, (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (pp. 264-271). Cambridge, MA: The MIT Press/Bradford Books.

Averill, J. R. (1979). Anger. In H. E. Howe and R. A. Dienstbier (Eds.), *Nebraska Symposium on Motivation 1978* (Vol. 26, pp. 1-80). Lincoln: University of Nebraska Press.

Averill, J. R. (1983). Studies on anger and aggression. Implications for theories of emotion. *American Psychologist*, 38, 1145-1160.

Baumeister, R. F. and Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, 117(3), 497-529.

Baumeister, R. F., Stillwell, A. M. and Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115(2), 243-267.

Becker, H. S. (1960). Notes on the Concept of Commitment. *American Journal of Sociology*, 66, 32-40.

Bloch, S. (1989). Émotion ressentie, émotion recréée. *Science et Vie, Hors série*, 168, 68-75.

Bloch, S., Orthous, P. and Santibanez-H, G. (1987). Effector Patterns of Basic Emotions: A Psychophysiological Method for Training Actors. *Journal of Social and Biological Structures*, 10, 1-19.

Boucher, J. D. (1983). Antecedents to Emotions Across Cultures. In S. H. Irvine and J. W. Berry (Eds.), *Human Assessment and Cultural Factors* (pp. 407-420). New York: Plenum.

Bowlby, J. (1969). *Attachment and Loss. Vol. 1: Attachment*. London: Penguin Books.

Bowlby, J. (1973). *Attachment and Loss. Vol. 2: Separation, Anxiety, and Anger*. London: Penguin Books.

Bowlby, J. (1980). *Attachment and Loss. Vol. 3: Loss, Sadness, and Depression*. London: Penguin Books.

Castelfranchi, C. (1990). Social power. A point missed in multi-agent, DAI and HCI. In Y. Demazeau, and J.-P. Müller, (Eds.), *Decentralized A.I.*, (pp. 49-62), North-Holland: Elsevier Science Publishers.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon Books.

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

Damasio, A. R. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. New York: Harcourt Brace and Company.

De Wall, F. (1996). *Good Natured. The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.

- Dimberg, U. (1982). Facial Reactions to Facial Expressions. *Psychophysiology*, 19(6), 643-647.
- Dongha, P. (1994). Toward a Formal Model of Commitment for Resource Bounded Agents. In M. J. Wooldridge and N. R. Jennings, (Eds.), *Intelligent Agents*, Proceedings of the ECAI-94 Workshop on Agent Theories, Architectures, and Languages, Lecture Notes on Artificial Intelligence, No. 890, (pp. 86-101), Berlin: Springer-Verlag.
- Eibl-Eibesfeldt, I. (1975). *Ethology: The Biology of Behavior*. (Second edition). New York: Holt, Rinehart and Winston.
- Ekman, P. (Ed.) (1982). *Emotion in the human face*. Cambridge, England: Cambridge University Press.
- Ekman, P. (1984). Expression and the Nature of Emotion. In K. R. Scherer and P. Ekman (Eds.), *Approaches to Emotion* (pp. 319-343). Hillsdale, N.J.: Erlbaum.
- Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist*, 48, 384-392.
- Ekman, P. and Davidson, R. J. (Eds.) (1994). *Can We Control our Emotions?* Chapter 7: of *The Nature of Emotion: Fundamental questions* (pp. 263-282). Oxford: Oxford University Press.
- Evans, D. (2002). *Emotion: The Science of Sentiment*. Oxford: Oxford University Press.
- Fikes, R. E. (1982). A Commitment-Based Framework for Describing Informal Cooperative Work. *Cognitive Science*, 6, 331-347.
- Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. New York: W. W. Norton and Company.
- Frijda, N. H. (1989). The Functions of Emotional Expression. In J. P. Forgas and J. M. Innes (Eds.), *Recent Advances in Social Psychology: An International Perspective* (pp. 205-217). Amsterdam: North-Holland.
- Gasser, L. (1991). Social Conceptions of Knowledge and Action: DAI Foundations and Open Systems Semantics. *Artificial Intelligence*, 47, 107-138.
- Gerson, E. H. (1976). On "Quality of Life". *American Sociological Review*, 41, 793-806.
- Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2), 161-178.
- Hatfield, E., Cacioppo, J. T and Rapson, R. L. (1994). *Emotional Contagion*. Cambridge, England: Cambridge University Press.
- Hoffman, H. S. (1974) Fear-Mediated Processes in the Context of Imprinting. In M. Lewis and L. A. Rosenblum (Eds.), *The Origins of Fear* (pp. 25-48). New York: John Wiley and Sons.

Jennings, N. R. (1993). Commitments and Conventions: The Foundation of Coordination in Multi-Agent Systems. *Knowledge Engineering Review*, 8, 223-250.

Kerr, N. L. and Kaufman-Gilliland, C. M. (1994). Communication, Commitment, and Cooperation in Social Dilemmas. *Journal of Personality and Social Psychology*, 66(3), 513-529.

Klennert, M. D., Campos, J. J., Sorce, J. F., Emde, R. N. and Svejda, M. (1983). Emotions as behavior regulators: Social referencing in infancy. In R. Plutchik and H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 2. Emotions in early development* (pp. 57-86). New York: Academic Press.

Krebs, D. (1987). The Challenge of Altruism in Biology and Psychology. In C. Crawford, M. Smith and D. Krebs, (Eds.), *Sociobiology and Psychology: Ideas, Issues and Applications*, (pp. 81-118). Hillsdale, NJ: Erlbaum.

LeDoux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.

Lemerise, E. A. and Dodge, A. D.(1993). The Development of Anger and Hostile Interactions. In M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 537-546). New York: The Guilford Press.

Levenson, R. W. and Ruef, A. M. (1992) Empathy: A Physiological Substrate. *Journal of Personality and Social Psychology*, 63(2), 234-246.

Lewis, M. (1993). Self-Conscious Emotions: Embarrassment, Pride, Shame, and Guilt. In M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 563-573). New York: The Guilford Press.

Lutz, T. (2001). *Crying: The Natural and Cultural History of Tears*. New York: W. W. Norton & Company.

MacLean, P. D. (1993). Cerebral Evolution of Emotion. In M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 67-83). New York: The Guilford Press.

Malatesta, C. Z. and Haviland, J. M. (1982). Learning Display Rules: The Socialization of Emotion Expression in Infancy. *Child Development*, 53, 991-1003.

Mead, G. H. (1934). *Mind, Self, and Society from the Standpoint of a Social Behaviorist*. Chicago: The University of Chicago Press.

Nesse, R. M. (Ed.) (2001). *Evolution and the Capacity for Commitment*. New York: Russell Sage Press.

Norman, D. A. (1980). Twelve Issues for Cognitive Science. *Cognitive Science*, 4, 1-32.

Oatley, K. and Johnson-Laird, P. N. (1996). The Communicative Theory of Emotions: Empirical Tests, Mental Models, and Implications for Social Interaction. *In* L. L. Martin and A. Tesser eds., *Striving and Feeling. Interactions among Goals, Affect and Self-Regulation*, 363-393. Hillsdale, NJ: Erlbaum.

Öhman, A. (1986). Face the Beast and Fear the Face: Animal and Social Fears as Prototypes for Evolutionary Analyses of Emotion. *Psychophysiology*, 23, 123-145.

Öhman, A. (1993). Fear and Anxiety as Emotional Phenomena: Clinical Phenomenology, Evolutionary Perspectives, and Information-Processing Mechanisms. *In* M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 511-536). New York: The Guilford Press.

Panksepp, J. (1991). Affective Neuroscience: A Conceptual Framework for the Neurobiological Study of Emotions. *In* K. T. Strongman (Ed.), *International Review of Studies on Emotion*, (Vol. 1, pp. 59-99). New York: John Wiley and Sons.

Provine, R. R. (2000). *Laughter: A Scientific Investigation*. New York: Penguin Books.

Rosenblum, L. A. and Alpert, S. (1974) Fear of Strangers and Specificity of Attachment in Monkeys. *In* M. Lewis and L. A. Rosenblum (Eds.), *The Origins of Fear* (pp. 165-193). New York: John Wiley and Sons.

Sagi, A. and Hoffman, M. L. (1976). Empathic Distress in the Newborn. *Developmental Psychology*, 12(2), 175-76.

Scherer, K. R. (Ed.) (1988). *Facets of Emotion: Recent Research*. Hillsdale, N.J.: Erlbaum.

Scherer, K. R., Wallbott, H. G. and Summerfield, A. B. (Eds.) (1986). *Experiencing Emotion: A Cross-Cultural Study*. Cambridge, England: Cambridge University Press.

Simon, H. (1967). Motivational and Emotional Controls of Cognition. *Psychological Review*, 74, 29-39.

Sloman, A. (1995). Architectures for Emotional Agents. Conference presented at the *Geneva Emotion Week*, Université de Genève, Geneva, April 8-13, 1995.

Tangney, J. P. (1995). Shame and Guilt in Interpersonal Relationships. *In* J. P. Tangney and K. W. Fischer (Eds.), *Self-Conscious Emotions. The Psychology of Shame, Guilt, Embarrassment, and Pride* (pp. 114-139). New York: The Guilford Press.

Tappolet, C. (2000). *Émotions et valeurs*. Paris: PUF, 296 p.

Toates, F. (1986). *Motivational Systems*. Cambridge: Cambridge University Press.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, 46, 35-57.

Vallerand, R. J. and Thill, E. E. (1993). Introduction au concept de motivation. In R. J. Vallerand and E. E. Thill (Eds.), *Introduction à la psychologie de la motivation* (pp. 3-39). Laval, Québec: Éditions Études Vivantes.