Artificial Emotions via Virtual Neuromodulators

Conference Paper · August 2014

CITATIONS
0

READS
94

2 authors:

Max Talanov
Kazan (Volga Region) Federal University
27 PUBLICATIONS **22** CITATIONS
SEE PROFILE

Alexander Toschev
Kazan (Volga Region) Federal University
12 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Artificial Emotions via Virtual Neuromodulators

Max Talanov[1] and Alexander Toschev[1]

Higher Institute of Information Technologies and Information Systems of Kazan Federal University

Abstract. How can we make machines actually feel emotions? Is there any option to make AI suffer, feel happiness, love, aggression, contempt, awe? Before we could find proper answer to this question we should be take in account several aspects: psychological as overall picture of low level and high level subjective emotions, neurophysiological as low level objective mechanism of brain response and cognitive architecture to put all these approaches in. We propose framework for emotional thinking implementation in computational spiking neurons based on Lövheim "Cube of emotions" [5] , Plutchik "Wheel of emotions" [10], Tomkins "Theory of affects" [4] and combine this theories in the cognitive architecture environment of "The emotion machine" of Marvin Minsky [7]. We propose mechanisms of emotions influence on computational processes via neuromodulation.
Keywords: AI, Cognitive architecture, Cognitive and Affective Modeling, Machine Thinking, Machine Understanding, Computing Emotions, Affective Computing, Neuromodulation, Neurotransmission, Model of Emotions, Model of Emotional Feedback Loops

1 Introduction

The biggest problem of artificial intelligence is that we don't know what natural intelligence is. The only example of intelligence that we observe today is human intelligence. Unfortunately there are no aliens so far who would be able to demonstrate non-human intelligence, and examples on Earth lack certain substance. Rosalind Picard in her article indicated: "There may exist a kind of alien intelligent living system, something weve never encountered, which achieves its intelligence without having anything like emotion. Although humans are the most marvellous example of intelligence we have, and we wish to build systems that are natural for humans to understand, these reasons for building human-like systems should not limit us to thinking only of human abilities." [9]

There are several domains that still remain unclear: creativity, intuition, insight and consciousness that prevent us from answering the question: "How David Lynch could create Mulholland Drive or Picasso could create Guernica?". We have started from possibly the first intelligent method emerged during evolution and from the most primal component of intelligence: the emotions.

Turing stated in his report "Intelligent Machinery" [16] that idea of intelligent machines "cannot be wholly ignored, because the idea of 'intelligence' is itself

emotional rather than mathematical". There is an interconnection between emotions and rational thought. Marvin Minsky in his book "The emotion machine" proposed that emotions are inseparable from thinking: "Emotional thinking: A flash of impatience or anger can cut through what seems like a hopelessly tangled knot. Each such 'emotional way to think' is a different way to deal with things, and some can increase your persistence or courage, while others can help you simplify things. In any case, after each such change, you may still want to pursue some similar goals, but now you'll see them from new points of view because each switch to a new Way to Think may initiate a large-scale cascade. Then, depending on how long those changes persist, you (or your friends) might recognize this as a change in your emotional state".

This article is dedicated to emotional aspects of human thinking and influence of emotions on computational processes in Marvin Minsky's cognitive architecture [7]. It could be considered as a base of computational emotional thinking framework and could be used in several domains like:

1. Advertisement
2. Emotional behavior simulations
3. Robotics
4. Intellectual assistants
5. Estimating human behaviour
6. Nursing software and robotics

2 Bases

Starting from the top we first reviewed several psychological models of emotions. Then we tried to understand the low level nature of emotions that brought us to neurochemical base of emotions. As we got the picture of human emotional processes we mapped them to cognitive architecture to gain AI basis. The first base is the evolutionary psychology theory of Plutchik [10]. We used Plutchik's main emotional process (feedback loops) and adapted it to Model of six [7] thinking levels of Marvin Minsky's cognitive architecture. We used "Wheel of emotions" [10] as a framework for subjective conscious and high level emotions. The second base is the theory of affects by Tomkins [4, 13, 14, 15]. We applied Tomkins theory of affects as the base for low level non-conscious emotions and low level appraisal. The neuro-physiological base is Lövheim theory of neuromodulatory base of emotions [5]. We used "Cube of emotions" as the main low level ("hardware") mechanism of emotional processes. All theories described above were mapped to Marvin Minsky's cognitive architecture described in his book "The emotion machine" [7].

3 Emotional Feedback Loops

Robert Plutchik created the three dimensional model [10] called "Wheel of emotions", that we used to describe subjective conscious emotions. There are eight

basic emotions grouped in pairs: joy - sorrow, anger - fear, acceptance - disgust, surprise - expectancy. We suppose that innate non-conscious affects and their appraisals could use different much quicker and simpler mechanisms than the mechanisms trained during lifetime: conscious emotions, appraisals and coping. Thus we use two models: affect theory for innate quick emotional reactions and "Wheel of emotions" for conscious low level and high level emotions. According to Plutchik theory basic emotions could be mixed like colors into high level emotions listed below:

1. Love = joy + acceptance
2. Submission = acceptance + fear
3. Awe = fear + surprise
4. Disapproval = surprise + sorrow
5. Remorse = sorrow + disgust
6. Contempt = disgust + anger
7. Aggressiveness = anger + expectancy
8. Optimism = expectancy + joy

Model presented above could be understood as the base for the subjective picture of emotions.

An example of emotional feedback loops [10] represented in four layers of Marvin Minsky's "Model of six" is presented on the figure 1.

We use four of the six layers just for the purpose of this example. Self-conscious reflections layer could influence emotions; for example evaluation of self as not progressing could cause sorrow or even depression, but it was not shown on the diagram.

We correspond instinctive reactions layer with non-conscious, innate, affective responses that mainly takes palace in: spinal cord, hypothalamus and amygdala. This way any stimulus is being processed first unconsciously; this is shown as affective appraisal rectangle. First stimulus triggers affective appraisal and affective appraisal triggers neuromodulation. Actually neuromodulation is not the result of appraisal, but non-conscious appraisal is accompanied by neuromodulation on the way from spinal cord to hypothalamus and amygdala. Using a concrete example: neuromodulation triggers an emotional state switch from serenity to fear. Affective appraisal triggers cognitive appraisal and reflective thinking. Cognitive appraisal in its turn initiates a deliberation process. Meanwhile second stimulus triggers second affective appraisal and its neuromodulation switches emotional state from fear to terror. Second affective appraisal triggers cognitive appraisal that in its turn initiates second appraisal deliberation. Then reflective thinking process estimating all activities in mind realizes that it's too emotional now and then stop all appraisal related processes and starts new coping oriented deliberation and switches emotional state via neuromodulation from terror to fear. Third appraisal deliberation selects cognitive reappraisal as coping strategy and this coping strategy is executed and switches emotional state back to serenity (via neuromodulation). In parallel to all cognitive and reflective process the second affective appraisal initiates non-conscious instinctive coping

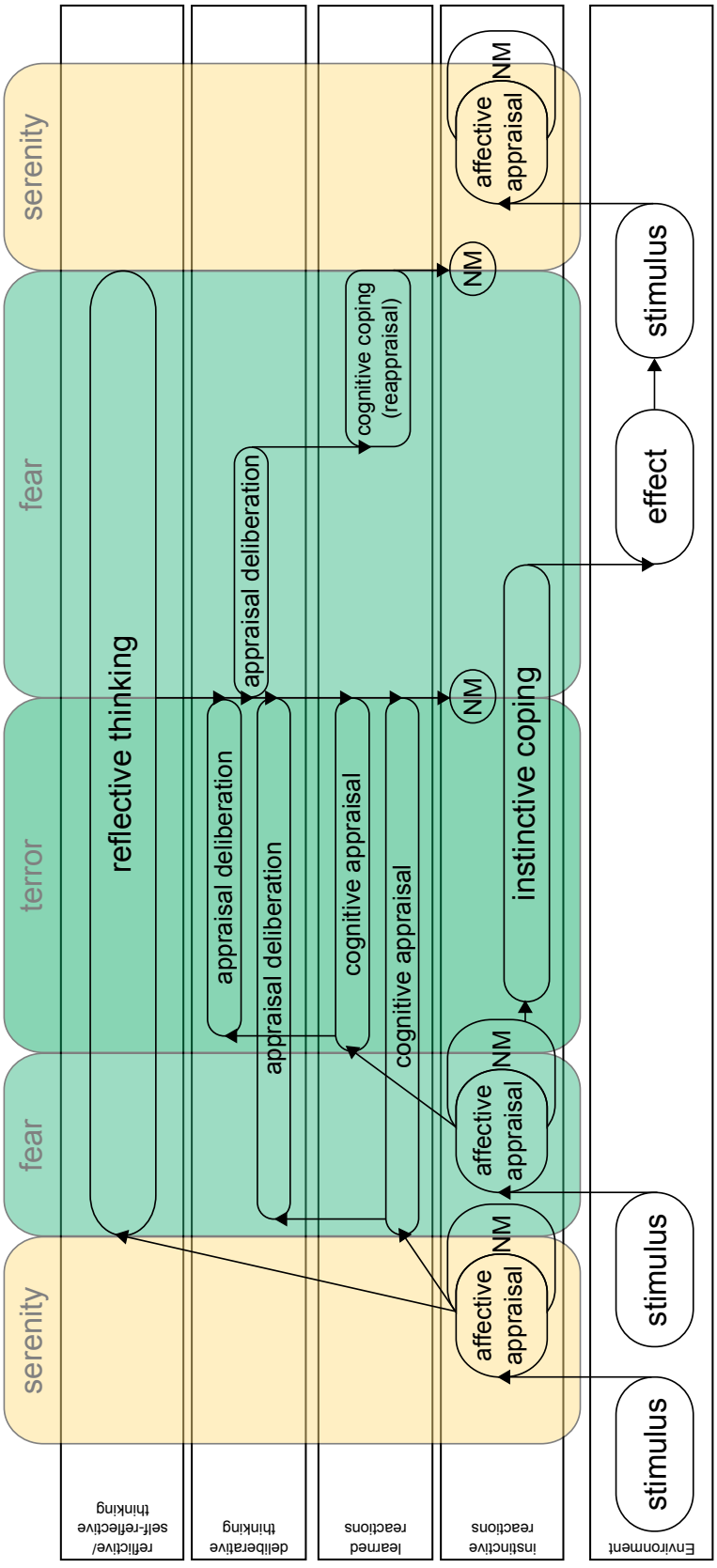

Fig. 1. Orchestra of emotions

strategy and it when applied created an effect over environment and this effect is been appraised again as stimulus.

4 Neuromodulatory Basis of Artificial Emotions

Hugo Lovheim in 2012 published his article "A new three-dimensional model for emotions and monoamine neurotransmitters" [5]. He described three dimensional model of emotions. Axes of the model are neuromodulators (monoamines): serotonin, dopamine, noradrenaline. Vertexes are affects from Tomkins affect theory [4]: enjoyment/joy, interest/excitement, surprise, anger/rage, disgust, distress/anguish, fear/terror, shame/humiliation. From our perspective this is the base of objective non-conscious emotional brain reaction to stimulus. On the other hand according to "Emotions: from brain to robot" [1] there are four following neuronal systems involved in emotional processing: spinal cord, hypothalamus, amygdala, frontal cortex, cingulate cortex. We roughly correspond non-conscious instinctive reactions layer of "model of six" [7] with spinal cord, hypothalamus and amygdala, while conscious processes and learned reactions, deliberative thinking, reflective thinking, self-reflective thinking, self-conscious reflections with frontal and cingulate cortex. This approach could be understood as subjective emotions to objective brain reaction mapping. This is fundamental for representation of emotional processes on computational system parameters.

5 Neuromodulators to Computing Parameters Mapping

Our understanding of the role of neuromodulators [5, 1] is represented in following mapping of neuromodulators to computing system parameters, the figure 2.

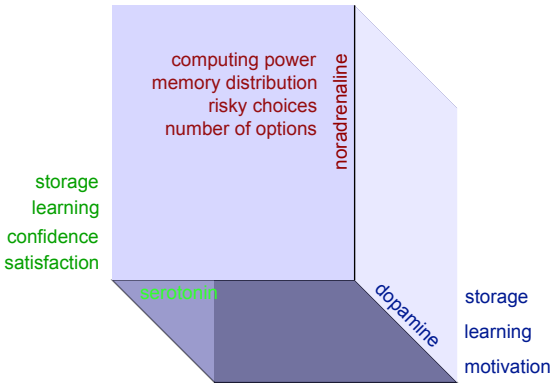

Fig. 2. Cube of parameters

5.1 Generic

Computing power: distribution and priority of parallel process or load balancing, is impacted by noradrenaline: the higher the level of noradrenaline is the more computing power must be concentrated on current activity (neuromodulator regulating attention).

Working memory(short term) distribution and concentration is impacted by noradrenaline (attention).

Learning is impacted by serotonin and dopamine: dopamine plays major role in activation of previously remembered patterns and serotonin in pattern generation.

Storage management (long term memory) is impacted by both by serotonin and dopamine, higher concentrations of both neuromodulators makes system better remember stimulus. In general, strong emotions generate more persistent memories.

5.2 Decision making

This decision making is done mainly in deliberation and learned reaction layers of model of six. Parameters: confidence, satisfaction, risky are used to highlight actions stored(remembered).

Confidence and satisfaction of the system is directly influenced by serotonin.

System is more *motivated* under the influence of dopamine.

System tends to choose *risky* actions under the influence of noradrenaline.

Noradrenaline makes system consider a smaller *number of options* in width and depth to be processed during deliberation.

This mapping is exhaustively described in "Computational emotional thinking and virtual neurotransmitters" [12]. It could be used as a low level ("hard-coded") model of emotional processes implemented in a spiking neuron model used to build a neural network and could be used as a basic framework for the emotion enabled systems [8].

6 Appraisal

Model described above operates in thinking processes environment, surrounded mainly by appraisal and coping processes [6]. We classify two types of appraisal processes mentioned above: non-conscious (quick, low level) and conscious (slower, high level) appraisals.

6.1 Non-conscious Appraisals

Non-conscious appraisals are associated with instinctive reactions layer of model of six, are actually performed in spinal cord, hypothalamus and amygdala and has innate nature formed during evolution. We used Tomkins theory of affects as a base for non-conscious emotional reactions [4]. The main criterion used for evaluation is activity level of CNS that could be steady, increasing, decreasing. Thus appraisal is described as following:

1. Quickest increase of brain activity triggers *surprise*, a bit slower increase - *fear/terror*, and most moderate - *interest/excitement*
2. Moderate steady CNS activity triggers *distress/anguish*, while high steady activity triggers *anger/rage*. It worth to note that the higher distress CNS activity is the easier is switch to anger. We could interpret this as follows: the longer the person is in distress state the easier he/she could be switched to anger state
3. Decrease of CNS activity is considered as relief and triggers *enjoyment/joy*
4. "*Disgust* NEGATIVE affect is inherently punishing and provides us some protection against eating poisonous or rotten food" [4]. We consider this affect as an unconditional rejection of inbound stimulus as something directly damaging the system. This could be understood as a low level hard-coded predicate to protect system
5. *Shame/humiliation* "affect is neither inherently punishing nor rewarding. It is like a computers reset button that rapidly clears the system and prepares us for whatever comes next... Without the innate affect shame-humiliation, we would not be motivated to take action when we are deprived of interesting and enjoyable things" [4]. This complex affect that appeared to be latest in the evolutional process of humans is triggered when a system was prevented to get new interesting information. Here, the social emotions of shame and humiliation are considered from the perspective of applicability outside a social context

6.2 Conscious Appraisal

Conscious appraisal has more complex self-emergent nature and is based on nurture and education of a child. We used "Appraisal Considered as a Process of Multilevel Sequential Checking" [11] as base for conscious appraisal process and derive patterns for Plutchik "Wheel of emotion" model of emotions. Please see for complete list of patterns [1]

Scherer uses 16 Stimulus Evaluation Checks (SEC) as basic blocks for whole appraisal processes. And overall process looks like the following sequence:

1. Relevance check including: novelty check, intrinsic pleasantness check, goal relevance check
2. Implication check including: causal attribution check, outcome probability check, discrepancy from expectation check, goal/need conduciveness check, urgency check
3. Coping potential including: control check, power check, adjustment check
4. Normative significance including: internal standards check, external standards check

[1] Comprehensive list of patterns available here: `https://github.com/development-team/2/blob/master/doc/emotions/affective\%20and\%20appraisal\%20aspects/appraisal_coping_high_level_emotions_aspects.md#comprehensive-secs-structure`

7 High Level and Low Level Emotions

Robert Plutchik [10] identifies 8 higher emotions, see Emotional feedback loops section. We suppose that these emotions should be represented as mixtures of neuromodulator levels of basic emotions. According to Plutchik higher emotions are combinations of basic emotions is the same manner as colors are mixed. Taking in account Lövheim model of neuromodulator relative levels and Gaussian nature of relation of subjective emotion perception to objective brain response [2, 17] we propose the following mapping of high level emotions on neuromodulator levels:

1. *G(Optimism Intensity)* = 1.0 serotonin + 1.0 dopamine + 0.5 noradrenaline;
2. *G(Love Intensity)* = 0.75 serotonin + 1.0 dopamine + 0.0 noradrenaline;
3. *G(Aggressiveness Intensity)* = 0.5 serotonin + 1.0 dopamine + 1.0 noradrenaline;
4. *G(Awe Intensity)* = 0.5 serotonin + 0.5 dopamine + 0.5 noradrenaline;
5. *G(Contempt Intensity)* = 0.5 serotonin + 0.5 dopamine + 0.5 noradrenaline;
6. *G(Disapproval Intensity)* = 0.5 serotonin + 0.0 dopamine + 1.0 noradrenaline;
7. *G(Submission Intensity)* = 0.25 serotonin + 1.0 dopamine + 0.0 noradrenaline;
8. *G(Remorse Intensity)* = 0.0 serotonin + 0.0 dopamine + 0.0 noradrenaline;

This could be understood as a basic mechanism for self-reflective thinking and self-conscious reflection layer influence on neuromodulation. Current model contains two high level emotions that have identical levels of neuromodulators: awe and contempt. We address both of them in the geometrical center of "Cube of emotions" [5] as intermediate states in between: fear and surprise for awe, and disgust and anger for contempt. One explanation is that the three dimensional model of Robert Plutchik "Wheel of emotions" contains only the basic emotions, all the high level, complex emotions are represented in 2 additional dimensions. Demonstrated approach produces a five dimensional model with axes: pleasantness (joy, sadness), attention (expectation, surprise), sensitivity (anger, fear), aptitude (acceptance, disgust) and strength of emotion [2] in contrast the neuromodulatory model of emotions of Lövheim contains only three dimensions which causes a lack of dimensionality in representing the five dimensional emotional model of Robert Plutchik. From neuroscience perspective some researchers mention the important role of opioids in emotions and it was not taken in account by Lövheim [1].

8 Conclusion

We created the synthetic theory of emotions based on four starting points: Robert Plutchick "Wheel of emotions" [10, 2] as high level subjective picture of emotions, Tomkins theory of affects [4] as low level objective framework of affects/emotions, Hugo Lövheim "Cube of emotions" [5] as objective neurophysiological mechanism of emotions and Marvin Minsky "The emotion machine"

[7] as cognitive architecture as implementation environment for all mechanisms listed above.

We used main assumption: there could be two frameworks of emotions: innate low level based on affects, and high level self-emerged during childhood nurture and education. These two frameworks should correspond and have mechanism to influence each other. This is done via the neuromodulatory base of emotions [5, 3]. Overall, emotional processes have following structure: stimulus non-conscious appraisal, neuromodulation (physiological emotional state switch), conscious appraisal with possible deliberation and possible coping strategy selection, coping strategy application over environment. As soon as coping, or some other behavior is applied over the environment, its state is appraised again as a new stimulus. This process is called feedback loop [10] and creates everlasting spiral process of emotions appraisal → neuromodulation (physiological impact) → coping. In the similar way high level thinking processes could influence the emotional state: for example reflective thinking could trigger a neuromodulation switching emotional state of a system and start/stop cognitive appraisal, deliberation. Thus neuromodulators are main actors of the objective brain response. We mapped their impact over computational system parameters:

1. Generic:
   (a) Computing power: noradrenaline
   (b) Memory distribution (attention): noradrenaline
   (c) Learning: serotonin, dopamine
   (d) Storage: serotonin, dopamine
2. Decision making/reward processing:
   (a) Confidence: serotonin
   (b) Satisfaction: serotonin
   (c) Motivation, wanting: dopamine
   (d) Risky choices inclination: noradrenaline
   (e) Number of options to process: noradrenaline

This mapping could be used as a main low level mechanism that build a bridge from the neurophysiological framework to the computational system and answers the question: "How emotions could influence computational system parameters?"

We briefly describe the appraisal and the coping processes closely related to emotional states and reactions. This creates overall closed and looped (spiral) system of emotional collaboration of a system with an environment and the emotional processing of internal psychological states.

The presented mapping completes the neuromodulation centric low level emotion framework, it could be used in a spiking neural network to implement emotional thinking phenomena in computational systems. We suppose it could be useful framework in several domains: from advertisement to nursing software and robotics.

9 Acknowledgment

Tero Keski-Valkama (MSc) for his constant support and review of our work and theories.

References

[1] Arbib, M., Fellous, J.M.: Emotions: from brain to robot. Trends in Cognitive Sciences 8(12), 554–559 (2004)
[2] Cambria, E., Hussain, A.: Sentic Computing. Techniques, Tools, and Applications. Springer (2012)
[3] Fellous, J.M.: Neuromodulatory basis of emotion. The Neuro-scientist 5, 283–294 (1999)
[4] Kelly, V.C.: A primer of affect psychology (2009)
[5] Lövheim, H.: A new three-dimensional model for emotions and monoamine neurotransmitters. Medical Hypotheses 78 78, 341–348 (2012)
[6] Marsella, S., Gratch, J.: Modeling coping behavior in virtual humans: Dont worry, be happy. In: Appears in the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (2003)
[7] Minsky, M.: The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon & Schuster (2007)
[8] Picard, R.W.: What does it mean for a computer to "have" emotions? In: Trappl, R., Petta, P., S., P. (eds.) Emotions in Humans and Artifacts (2001)
[9] Picard, R.W.: Affective computing: Challenges. International Journal of Human-Computer Studies 59, 55–64 (2003)
[10] Plutchik, R.: The nature of emotions. American Scientist 89(4), 344–350 (2001)
[11] Scherer, K.R.: Appraisal considered as a process of multilevel sequential checking. Appraisal Processes in Emotion: Theory, Methods, Research pp. pp. 92–120 (2001)
[12] Talanov, M., Toschev, A.: Computational emotional thinking and virtual neurotransmitters. International Journal of Synthetic Emotions (IJSE) (2014)
[13] Tomkins, S.: Affect imagery consciousness volume I the positive affects. New York: Springer Publishing Company (1962)
[14] Tomkins, S.: Affect imagery consciousness volume II the negative affects. New York: Springer Publishing Company (1963)
[15] Tomkins, S.: Affect imagery consciousness volume III the negative affects anger and fear. (1991)
[16] Turing, A.: Intelligent machinery. In: Copeland, B. (ed.) The Essential Turing: the ideas that gave birth to the Computer Age., p. 411. Oxford: Clarendon, 2004 (1948)
[17] Zeki, S., Romaya, J.P.: Neural correlates of hate. PLoS ONE 3(10) (2008)