The Emotional Mechanisms in NARS

Pei Wang[1], Max Talanov[2], and Patrick Hammer[3]

[1] Department of Computer and Information Sciences, Temple University, USA
pei.wang@temple.edu
[2] Higher Institute of Information Technologies, Kazan Federal University, Russia
dr.max@machine-cognition.org
[3] Institute for Software Technology, Graz University of Technology, Austria
patrickhammer9@hotmail.com

Abstract. This paper explains the conceptual design and experimental implementation of the components of NARS that are directly related to emotion. It is argued that emotion is necessary for an AGI system that has to work with insufficient knowledge and resources. This design is also compared to the other approaches in AGI research, as well as to the relevant aspects in the human brain.

1 Intelligence and Emotion

In biological systems, emotion is closely associated with drives like survival and reproduction, according to which decisions are made. On the contrary, a computer system has no biological drive, and the primary driving force are the tasks assigned to the system by the designer or the user. Consequently, mainstream AI study has ignored emotion, and this attitude is also justified by the traditional belief that emotion is basically a distraction in decision making, so should be avoided by a rational thinker.

In recent decades, the functions of emotion in cognition and thinking have been established by many works in cognitive science, and its necessity in computer systems has also been argued by researchers including Picard [6], Arbib [1] and Minsky [5]. More and more AGI models include emotion as a fundamental mechanism, as exemplified by the recent recent works [2, 10, 7, 8].

In this paper, the emotional mechanism in NARS, an AGI project, is briefly introduced and compared with those in the other AGI models, as well as the emotional mechanism in the human brain.

NARS (Non-Axiomatic Reasoning System) is a general-purpose AI designed in the framework of a reasoning system. Its conceptual cornerstone is the belief that intelligence is a form of adaptation and must obey the Assumption of Insufficient Knowledge and Resources (AIKR), meaning the system must manage its finite processing capability, open to novel tasks, respond to them in real time, and learn from its experience.

This belief implies that the system must be able to assess various objects in its external and internal environments with respect to its tasks, and treat them accordingly, so as to approach its overall objective. Such a need will require a mechanism that is similar to what we call "emotion" in human cognition, even though the objective of NARS is not to simulate the human brain in all details.

2 Desirability of Events

As a reasoning system, the overall objective of NARS is to successfully carry out its *tasks*, including absorbing new knowledge, answering questions, and achieving goals. For the current purpose, the last will be the focus of the discussion.

As defined in [16], a *goal* in NARS is an event (i.e., a statement with a time-dependent truth-value) to be realized by the system, that is, to have event E as a goal means the system has committed to do something to make E happen. For example, achieving the goal "Open Door #3" is represented within the system as a process to make event "Door #3 is open" true. But since in NARS "true" is a matter of degree, it actually means to make the truth-value of the statement as true as possible.

As an AGI, normally there are many goals in NARS at the same moment demanding to be realized. The system is "real-time" in the sense that these goals have time-requirements attached, with various levels of urgency. Since the system only has finite processing capability, the competition of resources among goals become inevitable. The goals can also contradict with each other in content. For example, one of them may want event "Door #3 is open" to be true, while another goal wants it to be false. As an open system, NARS does not require the consistency of the goals assigned to it by its designers and users, and does not guarantee the consistency of the derived goals.

Therefore, the system has to constantly manage conflicting or competing goals. To indicate the system's preference, each event has a *desire-value* associated, which is defined as the truth-value of the implication statement stating that the realization of the event will lead to a (unspecified) desired state. In this way, the desire-value of an event is defined as the truth-value of a statement, and therefore can be handled accordingly [16].

The desire-values of input goals are determined by the designers and users of the system, and these goals could be implanted in the system's memory or entered via the user interface. The derivation and revision of goals are carried out by NAL inference rules, which also calculate desire-values for derived goals. Derived goals are basically handled in the same way as input goals [15, 16].

Each time a new goal enters (either input or derived), the desire-value of the corresponding event is adjusted by the revision rule that merges the contribution of the new goal with the previous value. For example, if one goal requires "Door #3 is open" to be true while another one does the opposite, these desires are balanced against each other: the resulting desire-value of the event reflects the summary of the desires, and therefore resolves conflicting goals.

If the goal corresponds to an operation (an event the system can trigger whenever it decides to), the desire-value of the event in respect to the current moment is determined. If this desire-value exceeds the decision-threshold system parameter, the operation is executed when the goal gets selected. An operation can consist of simpler operations to be executed in a sequence or in parallel.

Additionally (also for not executable goals), the system does a "reality check" to see to what extent the desired goal is already fulfilled. Then the difference between desire and reality is used to adjust the *priority value* of the goal in resource competition.

The *satisfaction* value of each event is defined to be the compliment of the difference between its desire-value and its truth-value, where 1 means the system has got what it desires, and 0 the opposite.[4] If the event is an operation, the satisfaction value is obtained from the feedback of an execution, which indicates whether the execution was successful, as far as the system can tell. In this way, the satisfaction value of an event measures the system's appraisal of the current situation on the event.

3 Feelings of the System

As a reasoning system, NARS works by repeating an inference cycle, in each of them a step of inference is carried out. In such a step, an inference task is processed by interacting with a belief of the system, and the result may be a partial solution to the task, as well as new tasks. If the task is a goal, then the result can lead to the adjustment of the satisfaction of the corresponding event. If the goal is "Open Door #3" and now the door is actually opened, the system is satisfied on this matter; if the door is still not open after the system's effort, it is unsatisfied on this matter.

The system's appraisal of the current and recent situations in general is obtained by summarizing its satisfaction values on the recently noticed events into an overall (system-level) satisfaction value S. After each cycle, the satisfaction value S is updated to $rs + (1 - r)S$, where s is the satisfaction value of the task processed in the cycle, and r is a system parameter identifying the relative weight of the two factors. In general, r is between 0 and 1, and the larger it is, the larger is role played by the current satisfaction in the overall satisfaction. We let r be a constant, though it may also depend on other factors, such as the priority of the task just processed.

The current satisfaction value could enter the system's experience via a "mental operator" $feel$. A mental operation can be executed by NARS on its memory to carry out self-monitoring and self-control functions [16]. In this case, the operation $feel$(`SATISFIED`) generates an event reporting the current satisfaction value of the system. This operation could be explicitly invoked as a goal, or automatically triggered when the satisfaction is beyond the neutral zone (around 0.5, defined by a system parameter). Here the term `SATISFIED` indicates the

[4] To simplify the discussion, in the above description a truth-value (and desire-value) is used as if it is a single number. In NARS, it is actually a "frequency-confidence" pair, and the previous comparison is done on an "expectation" function of the truth-value, which combines the two factor into a single value. For details, see [16].

target of the feeling operator, which can also be invoked for other internal sensations, such as:

Alertness - summarizes the average difference between recently processed input and the corresponding anticipations, so as to roughly indicate the extent to which the current environment is familiar.

Busyness - summarizes the average priority values of the recently processed tasks.

Well-being - summarizes the overall measure of energy supply, I/O channel connection, device functioning, etc.

Whether the above feelings are also considered as "emotions" depends on whether the notion is used in a broad sense or a narrow sense, but no matter what they are called, they add "mental events" into the system's experience, which happen in its own "mind", and are directly perceived at an abstract level by the system.

4 Emotion in Concepts

A *concept* in NARS is a data structure that can be addressed by an internal ID called a "term", and contains the tasks and beliefs on the term. Consequently a goal is linked by all the concepts mentioned in the goal. For example, the goal "Open Door #3" is linked from the concepts for the terms *open*, *door*, and #3, respectively, as well as from the compound term for the event "Door #3 is open".

Concepts provide an intermediate level between the whole memory and the individual tasks (including goals) and beliefs. Because NARS uses a term logic, every inference step requires the premises (the task and the belief) to share a term, and consequently the inference can be considered as happening in the concept named by the shared term. This nature allows a concept to be a unit of processing in a distributed implementation of NARS.

According to the experience-grounded semantics of NARS, the meaning of a concept is determined by its contents, that is, the tasks and beliefs that show the relations of this concept with other concepts according to the experience of the system. Due to insufficient resources, tasks have priority values attached to indicate how often they will be accessed. When a concept is "fired", i.e., selected for processing, usually only part of its contents are involved.

Each concept also has a desire-value. As described above, if a concept is named by a term that is an event like "Door #3 is open", its desire-value comes from the related goals about this event. Now desire-value is also given to other terms, those that do not name events, such as *open* and *door*, even #3. Initially, these non-event terms have a neutral desire-value, so they are neither desired nor undesired. However, they may gradually become non-neutral by association with the system-level satisfaction value. The process is roughly like this: at the end of each inference cycle, the desire-value of the "fired concept" (i.e., within which the inference happen) is adjusted according to the current satisfaction value.

Roughly speaking, the concept is desirable if it associates with the satisfaction of the system.

Here we want to explore whether such a desire-value can explain emotions related to concepts which by its structure can not contain statements, as we think that it might be shown by the human mind. We also want to explore the effect of this type of emotion in self-control.

To bring this appraisal into the internal experience of the system, the feeling operator can be invoked with a term as argument, such as $feel(door)$, to generate an event indicating how much the system "likes" (or "dislikes") the term *door*. This operator can also be triggered by an extreme (high or low) desire-value in the concept.

Beside this "emotional indicator" in every concept, there are also special concepts whose meaning is especially emotional. The basic concepts in this group include feeling constants like `LIKE` and `SATISFIED`. These concepts provide the building blocks for the system's feelings and emotions.

Starting from the basic feelings, more complicated feelings can be built by combining them with the other concepts. For example, an event with the same desire-value may become different feelings when combined with other features, such as "it has happened" vs. "it will happen", "it is caused by the system itself" vs. "it is caused by someone else", "it is manageable" vs. "it is inevitable", etc. The new feelings are formed using the same composing rules as other compound terms, and their generation is experience-driven. For example, what "*happy*" means will be mostly learned, though still related to `SATISFIED`. These compound feelings may or may not correspond to human feelings.

5 Effects of Emotion

As described above, in NARS emotional information appears in two distinct forms:

- at "subconscious level" (outside experience), as desire-values and satisfaction values,
- at "conscious level" (inside experience), as events with emotional concepts.

Emotions in both forms contribute to the system's behaviors.

The emotional concepts in experience are processed as other concepts in inference. An important usage of them is to categorize situations from the system's viewpoint, as well as to develop strategies to deal with such situations. For instance, there may be many very different situations that can be categorized as "dangerous", so as to be handled with some common responses, such as "be careful". Without emotion, such categorizations may still be possible, though emotion provides a more natural and efficient approach.

The "emotion-specific" treatments mainly happen at the subconscious level, where the emotional information is used in various processes, such as

- The desire-values of concept is taken into account in attention allocation, where concepts with strong feeling (extreme desire-values) get more resources than those with weak feeling (neutral desire-values).

- After an inference step, if a goal is relatively satisfied, its priority is decreased accordingly, and the belief used in the step gets a higher priority, because of its usefulness.
- In the decision-making rule, the threshold for a decision is lower in high emotional situations, so as to allow quick responses.
- The overall satisfaction is used as feedback to adjust the desire-values of data items (concepts, tasks, beliefs), so that the ones associated with positive feeling are rewarded, and the ones associated with negative feeling punished. In this way, the system shows a "pleasure seeking" tendency, and its extent can be adjusted by a system parameter.
- When the system is "busy", tasks with low resource budget are simply ignored. The busyness value can be used in the priority–probability mapping to control the "degree of focus" of the system's attention.
- When the system is "alert", it spends more time to process new tasks in the input buffer, which means less time for the existing tasks in memory.
- When the system "does not feel well", it spends more time in the related self-maintenance tasks, which means less time for other tasks.

The above mechanisms have been mostly implemented, and are under testing and tuning, so at the moment have not produced profound results to be evaluated.

In the future, when NARS also needs to manage its own energy usage (such as in robots), emotion will play an important role in the decision of energy consumption. For example, in situations associated with high emotions, the system may spend more energy than in normal situations.

Another future usage of emotion is in communication with other systems, where emotion will play roles similar to those in human communications.

6 Comparison to Other Approaches

The current approaches to introducing emotion into computer systems actually have different objectives [6, 1]. The works in the field of *affective computing* mainly aim at the recognition and simulation of human emotions in human-computer interaction, while the works in AI/AGI mainly aim at giving computer their own emotions. For our purpose, the emotions in the computer system do not need to be similar to human emotion in details, but should serve the same cognitive functions.

The cognitive functions of emotion are usually divided into two major types, which can be called "internal and external" [1] or "intrapersonal and interpersonal" [10]. Either way, the former is in self-control according to experience, and the latter is in communication with other systems. On this topic, our position is to take the former as primary and basic, the latter as secondary and derivative. For this reason, the current work in NARS focuses on the control function of emotion, which is the appraisal of situation from the system's viewpoint, and the corresponding adjustments in behavior and resource allocation [1].

Traditional AI ignores emotion, since there is little need to choose among goals, which are assumed to be consistent, and within the system's capability.

Since NARS is designed under AIKR, the traditional assumption is no longer valid, and the system does need to handle conflicting and competing tasks, as well as to make quick and flexible responses to the environment in real time.

Though other AGI projects include emotional mechanisms for similar reasons, the concrete designs are all different. Here we only briefly compare NARS with MicroPsi [2] and Sigma [7].

MicroPsi grows out of a psychological theory, and therefore is closer to the reality of the human mind than NARS, which is identified with the human mind at a more abstract level. This difference shows in the motivational systems of them: MicroPsi has a motivational system with a set of built-in *drives*, and *goals* are situations where some need is satisfied. The basic drives meet physiological needs, social needs, and cognitive needs. On the contrary, NARS is a reasoning system, where a goal is an event to be realized, and in principle the system can be given any goal, as far as it can be expressed in the representation language of the system. For specific application, it is possible to implant certain "innate" goals or drives, though the design of the system does not assume any of them. Many "cognitive needs" of MicroPsi, such as those for *certainty*, *competence*, and *aesthetics*, are also pursued in NARS, but they are not explicitly expressed as goals, but implicitly embedded in the system's processing procedures and policies, so they can be referred to as "meta-goals" or "subconscious goals". Even with these differences, there are still similarities in these two systems, such as to pursue multiple goals at the same time, while giving them different relative priority.

The emotion mechanisms of both NARS and Sigma start at appraisal, where different situations have different levels of desirability. However, Sigma defines desirability by comparing a *state* with a goal state, while NARS does so on a *statements*, a partial description of states, as well as on a concept. Under AIKR, in NARS it cannot be assumed that the system can fully describe a state, either of the environment or of itself. Another difference is that the word "emotion" is used in a broader sense in Sigma than in NARS. For instance, the attention mechanism of NARS [16] is not considered as part of the emotional mechanism, as the latter is based on the appraisal of desirability and satisfaction only, though it is indeed closely related to the former.

In summary, in these AGI systems emotion plays similar roles. NARS differs from the other systems mainly because of its reasoning system framework and AIKR. Since all these systems are still far from fully developed, it is too early to tell which treatment of emotion works better.

7 Comparison to Human Emotions

The approach to emotions in NARS is biologically inspired and based on the functional similarity with mammalian basic emotions. We have inherited the neurobiological plausible approach from our previous works [9, 14], where validation and justification of the approach are provided. We are building the analogy between the influence of mammalian basic emotions or "affects" [11–13] on thinking and the influence of machine emotions on reasoning and decision-making processes of NARS. We reference the neurobiological nature of the emotions and identify the dopamine as main actor in the role of "wanting" or desire-values of NARS, described in the Section 2. Lövheim [4] emphasized the role of the dopamine in reward, reinforcement, and motivation. Arbib and Fellous [1] also indicated that dopamine key role in memory "linking emotion, cognition and consciousness". Serotonin "plays a crucial role in the modulation of aggression and in agonistic social interactions in many animals. ... serotonin has come to play a much broader role in cognitive and emotional regulation, particularly control of negative mood or affect" [1, 3], also it is main actor in self confidence, inner strength, and satisfaction [4]. This could be understood as neuromodulatory basis of the satisfaction value in the NARS system, described in the Section 3. Drawing the analogy between the noradrenaline influence on a brain and busyness of a system we could provide a set of emotional operations that build the basement for the machine affective states.

A modified "cube of emotions" is in Fig. 1, where the influence of virtual/machine neuromodulators on computational processes is added into a presentation of normal concentrations of neuromodulators.

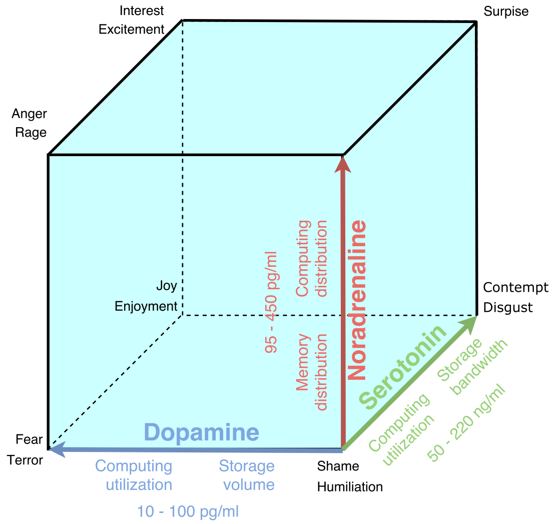

Fig. 1. The mapping of emotional states with neuromodulators levels and computational system parameters, based on [4] .

Computing utilization is a metric able to quantify how busy the processing resources of the system are. It can be expressed by the average value of all the single processing resources' utilization.

Computing distribution aims at quantifying the load balancing among processing resources. It can be expressed as the variance of single resources' utilization.

Memory distribution is associated with the amount of memory allocated to the processing resources. It can be quantified by the variance of the amount of memory per single resource.

Storage volume is an index related to the the amount of data and information used by the system.

Storage bandwidth quantifies the number of connections between resources, i.e. processing and data nodes.

Conceptually this work may lead to the integration between the neurobiologically plausible realistic neural networks (rNN) emotional simulations to computational lightweight reasoning systems applicable to real-time autonomous robotics. For example, a robotics system can enter experience into the system during a “day” phase, then this could be “played” into the rNN, similar to the dream playback in mammals. During the “night” phase, rNN could apply the realistic emotional processing. The results could be mapped through the levels of machine neuromodulators in NARS: serotonin, noradrenaline, dopamine, triggering the emotion-driven behavior.

8 Conclusions

This paper introduces the conceptual design of the emotion mechanism of NARS. We consider the main function of emotion as the appraisal of the external and internal entities and situations with respect to the system's tasks, so as to act accordingly, especially in decision making and resource allocation.

In NARS emotions are implemented not as an independent process or module, but are embedded in various places, and tightly entangled with the reasoning/learning processes in the system. The generation of emotion and feeling starts as desires for certain events, and the assessments to their satisfaction are summaries to the overall satisfaction of the system, and the association with this overall satisfaction determines the appraisal of concepts. Emotional information is taken into account in various places in the system, both consciously (i.e., expressed in the system's experience) and subconsciously (i.e., embedded in the system's built-in mechanisms).

The emotion of an AGI system will not be the same as human emotions, but since they play similar roles, some correspondence can be found between these two types of intelligence, mostly at psychological level, but may even at the neurobiological level to a certain extent. Though emotion may cause undesired consequences in decision making, it only means that the system must have mechanisms to regulate emotion, but not that high intelligence does not need emotion.

The emotional mechanism described in this paper has been mostly implemented in the current version of Open-NARS, an open source project.[5] The

[5] `https://github.com/opennars/opennars/wiki`

system is still under testing and tuning, so to show the function of emotion in the processing of complicated problems is still a future work.

Acknowledgments

Part of the work was performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

1. Arbib, M., Fellous, J.M.: Emotions: from brain to robot. Trends in Cognitive Sciences 8(12), 554–559 (2004)
2. Bach, J.: Modeling motivation and the emergence of affect in a cognitive agent. In: Wang, P., Goertzel, B. (eds.) Theoretical Foundations of Artificial General Intelligence, pp. 241–262. Atlantis Press, Paris (2012)
3. Fellous, J.M.: The neuromodulatory basis of emotion. The Neuro-scientist 5, 283–294 (1999)
4. Lövheim, H.: A new three-dimensional model for emotions and monoamine neurotransmitters. Medical Hypotheses 78, 341–348 (2012)
5. Minsky, M.: The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon & Schuster (2006)
6. Picard, R.W.: Affective Computing. MIT Press, Cambridge, Massachusetts (1997)
7. Rosenbloom, P.S., Gratch, J., Ustun, V.: Towards Emotion in Sigma: From Appraisal to Attention. In: Proceedings of AGI 2015. pp. 142 – 151. Springer International Publishing (2015)
8. Strannegård, C., Cirillo, S., Wessberg, J.: Emotional concept development. In: Proceedings of AGI 2015. pp. 362–372. Springer International Publishing (2015)
9. Talanov, M., Vallverdu, J., Distefano, S., Mazzara, M., Delhibabu, R.: Neuromodulating Cognitive Architecture: Towards Biomimetic Emotional AI. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications. pp. 587–592. IEEE (2015)
10. Thill, S., Lowe, R.: On the functional contributions of emotion mechanisms to (artificial) cognition and intelligence. In: Proceedings of AGI 2012. pp. 322–331. Springer International Publishing (2012)
11. Tomkins, S.: Affect imagery consciousness, Volume I: The positive affects. New York: Springer Publishing Company (1962)
12. Tomkins, S.: Affect imagery consciousness, Volume II: The negative affects. New York: Springer Publishing Company (1963)
13. Tomkins, S.: Affect imagery consciousness, Volume III: The negative affects: anger and fear. New York: Springer Publishing Company. (1991)
14. Vallverdú, J., Talanov, M., Distefano, S., Mazzara, M., Tchitchigin, A., Nurgaliev, I.: A cognitive architecture for the implementation of emotions in computing systems. Biologically Inspired Cognitive Architectures (2015)
15. Wang, P.: Motivation management in AGI systems. In: Proceedings of AGI 2012. pp. 352–361. Springer International Publishing (2012)
16. Wang, P.: Non-Axiomatic Logic: A Model of Intelligent Reasoning. World Scientific, Singapore (2013)