Functionalist Emotion Model in NARS

Xiang Li[1], Patrick Hammer[1](✉), Pei Wang[1](✉), and Hongling Xie[2](✉)

[1] Department of Computer and Information Sciences, Temple University, Philadelphia, USA
{xiangliAGI,tuh38867,pei.wang}@temple.edu

[2] Department of Psychology, Temple University, Philadelphia, USA
hongling.xie@temple.edu

Abstract. Emotions play a crucial role in different cognitive functions, such as action selection and decision-making processes. This paper describes a new appraisal model for the emotion mechanism of NARS, an AGI system. Different from the previous appraisal model where emotions are triggered by the specific context, the new appraisal evaluates the relations between the system and its goals, based on a new set of criteria, including desirability, belief, and anticipation. Our work focuses on the functions of emotions and how emotional reactions could help NARS to improve its various cognitive capacities.

Keywords: Artificial general intelligence · NARS · Emotion

1 Introduction

Emotion is intrinsic to human cognition and serves important functions in human cognition by providing critical information just like vision, hearing, touch, taste, and smell [3]. Emotion has only been the subject of study relatively recently, in the last centuries. The founder of American Psychology, William James, wrote one of the earliest treatments; "What is an Emotion?" [4] in 1884. Since then, many psychologists have tried to give a concrete definition of emotion. Many come to the same conclusion, exemplified by Dr. Joseph Ledoux's statement "*Unfortunately, ones of the most significant things ever said about emotion may be that everyone knows what it is until they are asked to define it*" [6]. This paper is not an attempt to produce such a definition. Instead it is meant to show an implementation of emotion within a specific cognitive architecture, NARS (Non-Axiomatic Reasoning System) [8,9].

The basic emotion mechanism of NARS has been discussed in previous publications [10,11]. NARS has a basic satisfaction-evaluation mechanism to indicate its "satisfaction" level, based on the degree of goals being achieved. In [10] an appraisal model has also been proposed where it generates emotions based on concrete events. For instance, when the system is hurt by an object, the system will subsequently react to this object with fear, and the object has to be clearly indicated. Recently, this design has been further aligned with the basic assumption of NARS that it should work with "*insufficient knowledge and resources*",

M. Iklé et al. (Eds.): AGI 2018, LNAI 10999, pp. 119–129, 2018.
https://doi.org/10.1007/978-3-319-97676-1_12

and also with the principle of an AGI system which is designed for general purposes. This paper will propose a new appraisal model in which emotions can be triggered by evaluating the events based on different criteria. Any events meeting those criteria will trigger corresponding emotions in NARS.

In the following, we will first review relevant concepts and proposals from psychology on emotion and its functions. We will then discuss the components related to the emotion mechanism of NARS, and finally, we will introduce the design decisions related to the new mechanism.

2 Psychological Studies on Human Emotion

Generally, the psychological study of emotion can be classified into two main schools: structuralism and functionalism. The structural perspective focuses on defining sets of features in an effort to build a taxonomy of basic emotions [1]. Understanding emotion is then the process of building a one-to-one relation between internal emotional states and observable artifacts: like facial expressions, autonomic responses, and changes in body chemistry. This perspective attempts to determine what areas of the brain are 'responsible' for a specific set of emotion, and how specific chemicals relate to their expression. Structuralists tend to neglect the study of intentionality.

In contrast, functionalists believe that emotion cannot be understood without understanding the motivations of the agent [1]. Their focus is on the impact of external and internal factors on a person's emotional state. Functionalists assert that facial expressions, gestures, and other indicators of emotion are also signals to other agents that affect the emotional state of others. Functionalism is a more pragmatic approach to studying machine emotion, making it the more appropriate perspective by which to investigate emotion as it may be applied to a computer system. This paper will adopt this perspective while outlining its approach to emotion.

Functionalists view emotion as relations between the external events and the internal goals of an agent and believe that emotion is synonymous with the 'significance' of the person-event transaction [1]. The significance of an event is related to how useful it is to the fulfillment of goals to the person. For the transaction to be significant at all, it must contribute to some goal.

To decide significance, Lazarus proposed an appraisal model which evaluates several factors related to goals and uses these appraisals to generate goal-related emotions [5]. Lazarus' model proceeds in three steps. First, the event is checked against current goals. If no goals are found to be related to the event, then there is no emotion triggered. Second, the congruence of the event with the goals is evaluated. This stage models how the event will impact the completion of the relevant goals, which can be either a positive or negative impact. The third stage is a type of ego-involvement, an involvement of one's self-esteem in the performance of a task or in an object, for example, ego-ideal, ego-identity, etc.

Denham [2] also designed an integrative model of three components (i.e., desire, state, and belief of certainty underline the cognitive process) for a child's experience of different emotions. The ***Prototype Approach*** describes the correlation between general types of events and specific emotions, each emotion is linked to common situations that cause it. For example, pleasurable stimuli or getting or doing something desired causes happiness. Anticipated harm or unfamiliar situations may cause fear, etc. Instead of encompassing emotional themes, the ***Event Structure Approach*** focuses on capturing the processes by which children come to experience different emotions. A child may experience fear if she realizes that it is very unlikely to maintain a desired state. In contrast, a child may experience anger if he realizes that some external conditions prevent him from achieving a desired state or avoiding an undesired state. The last approach is called ***Desire-Belief Approach***, and it describes how emotions may result from the consistency or discrepancy between one's desire or belief and the reality. A child who desires a gift feels happy if he actually gets one; in contrast, a child who believes Mom is sleeping in the bedroom may feel surprised when she finds nobody there. Based on these three components, Denham proposed an integrative model that encompasses both the process and the content of a child's reasoning that leads to different emotions. Table 1 shows the model for Happiness, Sadness, Anger, and Fear.

Table 1. Integrative model of Happiness, Sadness, Anger and Fear, by Susanne A. Denham

Desire	Want	Want	Want	Not want
State	Have	Not have	Not have	Have
Belief of certainty	Yes	Never	Can reinstate	Likely
Emotion	Happiness	Sadness	Anger	Fear

3 NARS Overview

NARS is an AGI built in the framework of a reasoning system and founded on the belief that "Intelligence" can be defined as *the ability for a system to adapt to its environment and to work with insufficient knowledge and resources.* This is captured by the acronym AIKR; Assumption of Insufficient Knowledge and Resources. AIKR and the NARS system are discussed in many publications, including two books [8,9]. This section will only cover the aspects of NARS most relevant to the current discussion.

NARS makes use of a formal language, "Narsese", for its knowledge representation, and this language is defined using a formal grammar in [9]. The system's logic is developed from the traditional "term logic". Statements in this logic have the form *subject-copula-predicate*. The smallest element that can be used as one of these components is referred to as a "term".

The most basic statement in Narsese is the *inheritance statement*, with the format "$S \rightarrow P$", where S is the subject term, and P is the predicate term. The "$\rightarrow$" is the *inheritance* copula, defined in ideal situations as a reflexive and transitive relation from one term to another term. The intuitive meaning of "$S \rightarrow P$" is "S is a special case of P" and "P is a general case of S". For example, the statement "$robin \rightarrow bird$" intuitively means "Robin is a type of bird".

At an atomic level, terms are simply sentences formed over a finite alphabet. In this article, terms are given a semantic meaning that is easily understood by a human reader. Terms like *wolf* or *animal* have some suggested meaning to the reader, but this is not required. Aside from atomic terms, Narsese also includes *compound terms* of various types. A compound term $(con, C_1, C_2, ..., C_n)$ is formed by a term connector, *con*, and one or more component terms $(C_1, C_2, ..., C_n)$. The term connector is a logical constant with predefined meaning in the system. Major types of compound terms in Narsese include

- **Sets:** Term $\{Tom, Jerry\}$ is an *extensional set* specified by enumerating its instances; term $[small, yellow]$ is an *intensional set* specified by enumerating its properties.
- **Products and images:** The relation "Tom is the uncle of Jerry" is represented as "$(\{Tom\} \times \{Jerry\}) \rightarrow uncle\text{-}of$", "$\{Tom\} \rightarrow (uncle\text{-}of \,/\, \diamond\ \{Jerry\})$", and "$\{Jerry\} \rightarrow (uncle\text{-}of \,/\, \{Tom\} \diamond)$", equivalently.
- **Statement:** "Tom knows snow is white" can be represented as a *higher-order statement* "$\{Tom\} \rightarrow (know \,/\, \diamond\ \{snow \rightarrow [white]\})$", where the statement "$snow \rightarrow [white]$" is used as a term.
- **Compound statements:** Statements can be combined using term connectors for disjunction('$\vee$'), conjunction('$\wedge$'), and negation('$\neg$'), which are intuitively similar to those in propositional logic, but not defined using truth-tables [8].

Several term connectors can be extended to take more than two component terms. The connector is then written before the components rather than between them, such as $(\times\{Tom\}\,\{Jerry\})$.

Beside the *inheritance* copula ('$\rightarrow$', "is a type of"), Narsese also includes three other basic copulas: *similarity* ('$\leftrightarrow$', "is similar to"), *implication* ('$\Rightarrow$', "if-then"), and *equivalence* ('$\Leftrightarrow$', "if-and-only-if"). The last two copulas are "higher order", meant to be applied to statements themselves.

In NARS, an *event* is a statement with temporal attributes. Based on their occurrence order, two events E_1 and E_2 may have one of the following basic temporal relations:

- E_1 happens before E_2
- E_1 happens after E_2
- E_1 happens when E_2 happen

Temporal statements are formed by combining the above basic temporal relations with some logical relations indicated by the term connectors and copulas. For example, the implication statement "$E_1 \Rightarrow E_2$" has three temporal versions, corresponding to the three temporal order relations:

- $E_1 \mathrel{/\!\!\Rightarrow} E_2$ (E_1 happens before E_2 and implies it)
- $E_1 \mathrel{\backslash\!\!\Rightarrow} E_2$ (E_1 happens after E_2 and implies it)
- $E_1 \mathrel{|\!\!\Rightarrow} E_2$ (E_1 happens when E_2 is happening and implies it)

These statements can be interpreted as a 'third-person' view of the statements or events. Narsese can also provide a 'first-person' view of events, describing things that the reasoning system may directly realize themselves. These special events are referred to as *operations*. These operations are tied to executable commands or procedures built or plugged into the system.

Formally, an operation is an application of an operator on a list of arguments, written as $op(a_1, \ldots, a_n)$ where op is the operator, and $a_1, ..., a_n$ is a list of arguments. Such an operation is interpreted logically as statement "$(\times\, \{SELF\}\, \{a_1\}\, \ldots\, \{a_n\}) \rightarrow op$", where $SELF$ is a special term indicating the system itself, and op is a term that has a procedural interpretation. For instance, if we want to describe an event "The system is holding key_001", the statement can be expressed as "$(\times\{SELF\}\ \{key\_001\}) \rightarrow hold$".

There are three types of sentences in Narsese:

- A **judgment** is a statement with a truth value, and represents a piece of new knowledge that system needs to learn or consider. For example, "$\langle robin \rightarrow bird\rangle.$" with a truth-value makes the system to absorb this conceptual relation, together with its implications, into the system's beliefs. More details about the truth value can be found in [8].
- A **goal** is a statement to be realized by executing some operations. For example, "$\langle(\times\, \{SELF\}\, \{door\_001\}) \rightarrow open\rangle!$" means the system should open the $door\_001$ or make sure that $door\_001$ is opened. Each goal associates with a "desire-value" indicating the extent to which the system hopes for a situation where the statement is true. More details about the desire value can be found in [8], too.
- A **question** is a statement without a truth-value or desire-value, and represents a query to be answered according to the system's beliefs or goals. For example, if the system has a belief "$robin \rightarrow bird$" (with a truth-value), it can be used to answer question "$\langle robin \rightarrow bird\rangle?$" by reporting the truth-value, as well as to answer the question "$\langle robin \rightarrow ?x\rangle?$" by reporting the truth-value together with the term *bird*, as it is in the intension of *robin*. Similarly, the same belief can also be used to answer question "$\langle ?y \rightarrow bird\rangle?$" by reporting the truth-value together with the term *robin*.

NARS' beliefs about itself start with its built-in operations. As mentioned above, the operation $op(a_1, \ldots, a_n)$ corresponds to a relation that the system can establish between itself and the arguments, as expressed by the statement

"$(\times\, \{SELF\}\, \{a_1\}\, \ldots\, \{a_n\}) \rightarrow op$" (where the subject term is a *product* term written in the prefix format), since it specifies a relation among the arguments plus the system identified by the special term $SELF$.

An operation may be completely executed by the actuator of the host system (e.g., a NARS-controlled robot raises a hand or moves forward), or partly by another coupled system or device (e.g., a NARS-controlled robot pushes a button or issues a command to another system). NARS has an interface for such "external" operations to be registered. Additionally there are "internal" or "mental" operations that can cause changes within the system.

In general, mental operations supplement and influence the automatic control mechanism, and let certain actions be taken as the consequence of inference. Mental operations contribute to the system's self-concept by telling the system what is going on in its mind, and allow the system to control its own thinking process to a certain extent. For instance, the system can explicitly plan its processing of a certain type of task. After the design and implementation phases, the system needs to learn how to properly use its mental operations, just like it needs to learn about the other (external) operations.

With regard to the current discussion, there are several important mental operations:

– **believe** generates a belief about a certain statement where the premises are not those covered by the existing inference rules. For instance, such a belief can be derived from a goal or a question.
– **want** is used to increase the desire-value of a statement, also in ways beyond what have been covered by the goal-derivation rules. When the desire-value exceeds a certain threshold, a goal is generated, and the event is recorded in the system's internal experience.
– **anticipate** allows NARS to predict the observation of an event. If the predicted event does not occur in time, the system will notice and more attention will given to the involved concepts. Additionally, a "disappointment' event can be generated, allowing the system to draw conclusions from the absence of a predicted event.

4 Appraisal Model in NARS

We take the position that emotions arise from cognitions regarding the outside world [7], through an appraisal process. Appraisal starts with extracting relevant information from its experience. No matter it is to a human being or to an intelligent agent, information about the outside world is not always prepared and waiting for the agent to receive. NARS is designed to handle such conditions, in a manner similar to a human agent.

To implement emotion, we need to concern ourselves with what kind of events might trigger emotion and apply the appraisal framework to these events. According to the previous discussion, the events that may trigger emotions are events that related to goals. This corresponds to the first stage of Lazarus's appraisal theory [5] which stated that only events related to goals would trigger emotions. If there is no goal concerning the event, then no emotion will be triggered.

The second stage of Lazarus's appraisal theory is to evaluate the congruence of the event with the agent's current goals. The result of this stage determines whether the emotion is positive or negative. In NARS, this evaluation is carried out by comparing the two measurements on an event: its desire-value and truth-value. If we assign binary values to desire-value (True for want, False for not want), truth-value (True for have, False for not have), and satisfaction value (True for positivity, False for negativity), the latter behaves exactly like the logic of an XNOR gate with the former two as inputs, as shown in Table 2.

Table 2. The relation among desire-value, truth-value, and satisfy-value

Desire value	Truth value	Satisfaction
Want	Have	Positive
Want	Not have	Negative
Not want	Have	Negative
Not want	Not have	Positive

The third stage of Lazarus's model, estimating the types of ego-involvement, is not considered in our appraisal model. Doing so would involve listing all possible types of ego-involvement, which is incongruent with the idea of being general-purpose, as well as with the situations where the same event may carry different meanings to a machine than to a human being. Also, we can interpret Denham's Belief of Certainty model as an uncertainty about the state, this can be represented by the tense of the event. If it is "yes", or "never", then the event should have already happened. If it is "likely to happen", it means the event has not happened but has a possibility to happen in the future. Our treatment addresses the implicit distinction of tense in the third row. However, we do not make a binary distinction based on the certainty as we believe the transition from Anxiety to Fear is gradual.

The following is how we defined criteria for some emotions:

- **Fear:** An undesired event is anticipated to happen
- **Happiness:** An desired event is believed as already happened (high satisfaction)
- **Sadness:** An undesired event is believed as already happened (low satisfaction)

The current design of NARS does not include a mechanism to simulate the physiological changes that accompany human emotional experiences such as changes in voltage or temperature. However, this does not mean that the system cannot "feel" anything. The basic feeling mechanism in NARS has been introduced in [11]. Feelings of emotions are realized by the *feel* operator through accessing the relevant sensors and returning the sensed state into the system's inner experience.

For example, if the system feels fear, the experience can be represented as

```
(^feel, {SELF}, fear).
```

Once some event matches the criteria of being afraid, the system will feel fear, implemented by the *feel* operator and the system will report the emotion it is currently feeling when such a question is asked.

So far, we have introduced how NARS generates different emotions by the new appraisal model, and how it feels these emotions. Examples of how NARS takes actions triggered by these emotions will be displayed in the next section.

In humans, emotions often improve their cognitive faculties. Allowing quicker reaction times, aiding in planning, and improving communication are all the important functions of an emotion mechanism. Similar benefits are gained by NARS.

Emotion helps the system summarize its experience at a more abstract level. For example; suppose an agent were to have some understanding of the concept of a 'beast' and had experienced instances that made the agent feel fear (e.g., seeing a wolf). This abstract notion of fear, as related to the concept of *beast*, may provide additional information about how to act when encountering a new member of the class *beast* (e.g. a bear). The emotional parameter offers a direct, concise way to encode information critical to the survival of the agent. Such summaries reduce the need for extra concepts to encode the relation between the source of the emotion and its outcomes.

Emotions can also decrease the time needed to respond to certain events. Emotional constructs like fear can facilitate a quicker response. Assuming that the emotions of the system are tied to more 'intrinsic' fundamental elements of the system, fear could be implemented as a response to anything that would impede the working of these fundamental systems. If fear is experienced, it is due to the agent being exposed to a situation or event that negatively impacted such systems. Any future event that produces the same interpretation could be responded to quickly.

5 Example

In the following, we illustrate an example using the Open-NARS implementation of NARS to show how emotions raise from the evaluations of events, and how emotions trigger actions. Due to the space limitation, we cannot explain the details of the representation, which can be found in [8,9].

```
=========================Happiness=========================

//Meaning of the statement: If something is wanted by SELF,
//and SELF's belief agrees with the case, SELF feels Happy

//1. #1 is a dependent variable which represents a certain
//   unspecified term under a given restriction. It can be
//   either an object or an event

//2. (^want, {SELF}, #1, TRUE) represents a mental operation
//   means something is desired by SELF; TRUE indicates the
//   truth value of this mental operation, where #1 is desired,
//   otherwise, use FALSE

//3. (^believe, {SELF}, #1,TRUE) means SELF's belief agree
//   with #1, if #1 represents an event, it indicates that
//   #1 has already happened.

//4. (^feel, {SELF}, happy) implements feel operator and
//   indicates the feeling of SELF being happy

//5. && is a term connecter, it connects the follow
//   term by meaning ``and''

Input: <(&&, (^want, {SELF}, #1, TRUE), (^believe, {SELF},
#1,TRUE)) =|> (^feel, {SELF}, happy)>.

//SELF has a goal which is not being hurt, ``--'' is the negation
//of the statement

Input: (--,<{SELF} --> hurt>)!

//SELF is not getting hurt, :|: represents the tense ``present''
//means SELF is not getting hurt right now

Input: (--,<{SELF} --> hurt>). :|:

//What do you feel?
//This statement is a question, and it corresponding to
//(^feel, {SELF}, happy) where ``?what'' at the position of the
//emotion

Input: (^feel,{SELF},?what)?

//SELF feels Happy, the reason why it feels happy is because
//SELF doesn't want to get hurt (generated by goal), and SELF
//is not getting hurt (generated by belief).

Answer: (^feel,{SELF}, happy).
```

```
===========================Fear================================

//If something is wanted by SELF, and SELF anticipates the
//opposite to happen, SELF feels fear
Input:<(&&, (^want, {SELF}, #1, FALSE), (^anticipate, {SELF},
#1)) =|> (^feel, {SELF}, fear)>.

//At the same time when SELF feels fear, it generate an
//motivation which to run away, run is also an operator in NARS

Input: <(^feel,{SELF}, fear) =|> <(*, {SELF},
<(*, {SELF}) --> ^run>) --> ^want>>.

//SELF doesn't want to be hurt

Input: (--,<{SELF} --> hurt>)!

//If wolf is getting close to SELF, SELF will get hurt
//&/ is another term connector representing the relation between
//two terms is ‘‘and’’, also the latter happens after the former.
//42 represents inference steps, it means, when wolf start
//getting close to SELF, after 42 steps, the SELF will get
//hurt. The number is not fixed, it can be any integer.

Input: <(&/,<(*, {SELF}, wolf) --> close_to>,+42) =/>
  <{SELF} --> [hurt]>>.

//Wolf is getting close to self

Input: <(*, {SELF}, wolf) --> close_to>. :|:

//Result: SELF takes the action run, based on the knowledge
//where SELF runs when it feels fear, SELF also feels the emotion
//fear
EXECUTE (^run,{SELF})
```

6 Conclusion

In this paper we introduced several new emotions to the NARS framework; *fear*, *sadness*, *happiness*, and *disappointment*. These additions were partially motivated by the descriptions of emotion provided by [2,5], but also came from the need to improve the control mechanism of NARS. In human, emotion results from a combined evaluation of belief, desire, and anticipation. This paper outlined how analogous processes in NARS can work to interpret a combination of parameters as an effective emotion.

Our results show that emotions could make the system to take actions in various situations. Emotion provides information that the system can use by offering a concise summary of the system's past experience with respect to its emotional state. In addition, it provides an additional mode of communication (system-to-system, human-to-system). Such functionality is important for a general-purpose intelligence system, especially if it should operate under the assumption of insufficient knowledge and resources (AIKR).

This work is not intended to produce an AGI with an emotional system just like that of a human being. Instead, the intent is to draw analogies between the human emotional system and components of a reasoning system in such a way that the reasoning system is improved. The basic emotions described in this paper provide a first step in establishing an effective emotional mechanism within NARS. Future work will be directed at building a richer experience for NARS. For instance, new emotional states, like regret, may provide additional feedback for the system to learn about prior errors.

References

1. Campos, J.J., Mumme, D.L., Kermoian, R., Campos, R.G.: A functionalist perspective on the nature of emotion. In: The Development of Emotion Regulation: Biological and Behavioral Considerations, vol. 59, pp. 284–303 (1994)
2. Denham, S.A.: Emotional Development in Young Children. Guilford Publications, New York City (1998)
3. Dijk, S.V.: Calming the Emotional Storm: Using Dialectical Behavior Therapy Skills to Manage Your Emotions and Balance Your Life. New Harbinger Publications, Oakland (2012)
4. James, W.: What is an emotion. Mind **9**, 188–205 (1884)
5. Lazarus, R.S.: Emotion and Adaptation. Oxford University Press, Oxford (1991)
6. Ledoux, J.: The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Simon and Schuster, New Jersey (1998)
7. Steunebrink, B.: The logical structure of emotions. Ph.D. thesis, Utrecht University Repository (2010)
8. Wang, P.: Rigid Flexibility: The Logic of Intelligence. Springer, Dordrecht (2006)
9. Wang, P.: Non-Axiomatic Logic: A Model of Intelligent Reasoning. World Scientific, Singapore (2013)
10. Wang, P., Li, X., Hammer, P.: Self in NARS, an AGI System. In: Frontier in Robotics and AI 5 (2018)
11. Wang, P., Talanov, M., Hammer, P.: The emotional mechanisms in NARS. In: Steunebrink, B., Wang, P., Goertzel, B. (eds.) AGI -2016. LNCS (LNAI), vol. 9782, pp. 150–159. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41649-6_15