

Article

Smart Sensing: An Info-Structural Model of Cognition for Non-Interacting Agents

Gerardo Iovane ¹, Iana Fominska ², Riccardo Emanuele Landi ^{3,*} and Francesco Terrone ⁴¹ Department of Computer Science, University of Salerno, 84084 Fisciano, Italy; giovane@unisa.it² Faculty of Nursing, Vinnytsya Medical College “D. K. Zabolotny”, 21037 Vinnytsya, Ukraine; janafominska5@gmail.com³ Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milano, Italy⁴ Direzione, Sidelmed S.p.A., 84085 Mercato San Severino, Italy; direzione@sidelmed.it

* Correspondence: riccardoemanuele.landi@mail.polimi.it

Received: 27 August 2020; Accepted: 7 October 2020; Published: 15 October 2020



Abstract: This study explores an info-structural model of cognition for non-interacting agents affected by human sensation, perception, emotion, and affection. We do not analyze the neuroscientific or psychological debate concerning the human mind working, but we underline the importance of modeling the above cognitive levels when designing artificial intelligence agents. Our aim was to start a reflection on the computational reproduction of intelligence, providing a methodological approach through which the aforementioned human factors in autonomous systems are enhanced. The presented model must be intended as part of a larger one, which also includes concepts of attention, awareness, and consciousness. Experiments have been performed by providing visual stimuli to the proposed model, coupling the emotion cognitive level with a supervised learner to produce artificial emotional activity. For this purpose, performances with Random Forest and XGBoost have been compared and, with the latter algorithm, 85% accuracy and 92% coherency over predefined emotional episodes have been achieved. The model has also been tested on emotional episodes that are different from those related to the training phase, and a decrease in accuracy and coherency has been observed. Furthermore, by decreasing the weight related to the emotion cognitive instances, the model reaches the same performances recorded during the evaluation phase. In general, the framework achieves a first emotional generalization responsiveness of 94% and presents an approximately constant relative frequency related to the agent's displayed emotions.

Keywords: affective computing; sensation; perception; emotion; affection; decision support systems; artificial consciousness

1. Introduction

Affective computing supports artificial intelligence by designing technologies that allow computational systems to recognize and elaborate human emotions and affections [1], enriching Decision Support Systems' features in making decisions, from emotional classification [2] to biometry [3]. Up to now, the research community has gained information about human emotions and affections [4–7]. Current technologies try to go deeper in recognizing “profound” human features [8–10], rather than just recognizing shapes, objects, or faces. Applied to autonomous agents, brain- and consciousness-related scientific research continuously spread towards a wide range of study fields, such as those related to psychiatric disorders [11], Neuroscience, and Cognitive Psychology. It is interesting, though, to examine in depth how the human cognitive scale leads to intelligence and how it can be translated into technology. Computational

models of cognition are of great utility when used to simulate cognitive processes, providing testbeds for cognitive scientists to evaluate their hypotheses [12]. Proposing formal models of cognition does not represent a reductionism of human mind description, but introduces new frontiers of comprehension.

In this paper, we do not want to enter the philosophical, psychological, neuroscientific, or medical debate about brain functioning. We limit our work to the analysis of a selected set of human cognitive levels, with the aim of translating mathematically their hierarchy into simple computable models. By means of empirical experience and scientific suggestions, we intend to contribute the widening of the way artificial intelligence is conceived; we think that the computational reproduction of human intelligence should not be based only on a single learning layer—for example, a model designed with just one neural network that classifies sensory inputs—but also on further levels of cognition related to those we think are the most important human factors. It is necessary to pursue the building of new paradigms of artificial intelligence that can be compared with human attitudes. Classification and regression algorithms, when exclusively based on sensory inputs, represent a conceptual reduction with respect to human intelligence, since human beings do not just recognize shapes or facial expressions. To be considered as “intelligent”, an agent should take decisions not only by processing sensory samples, but also by taking into account cognitive levels like those of emotion and consciousness.

Is it sufficient to consider just a neural network as intelligent? Do we need to design further levels of learning, in order to simulate reliable human cognition? The answer lies in the deepening of human cognition comprehension.

Although this paper mainly impacts Smart Sensing, it is prodromal to the rational integration of consciousness in solutions with artificial intelligence. In fact, already in the present work, the modeling of artificial sensations, perceptions, emotions, and affections is analyzed in a framework that sees attention, awareness, and cognition at the sensory-cognitive stages. These last stages will be discussed in the next paper, which will allow us to show the complete methodology on an artificial consciousness interposed between sensing and artificial intelligence. Such a vision is absolutely new in the international scenario since, without dealing with religious, ethical, or psychological issues, it allows us to strengthen cognition in the context of artificial thinking by enforcing human ability to create more sophisticated artificial intelligence. In the last 10 years, there have been many works in the context of Smart Sensing, especially if we consider the applications implemented with Industry 4.0 and with the Internet of Things [13,14]. The European Commission has promoted and financed different initiatives favoring human-machine interaction thanks to emotional involvement. For example, with the 7th Framework Programme, the ALICE (Adaptive Learning via Intuitive/Interactive, Collaborative and Emotional systems) project showed the effect of a type of learning that increased the level of attention of the learners, thanks to an analysis of learning styles and emotional involvement. With Horizon 2020, not only emotions but also affections have become the center of attention in different projects with international impact. They concern not only learning assisted by artificial intelligence technologies, but also automotive, online trading, and more generally computational finance, customer profiling in digital marketing, etc. This work frames Smart Sensing, specifically perceptions, emotions, and affections, as a bridge towards artificial consciousness to solve the gap between sensing and artificial intelligence, enriching the latter with elements that allow us to take another step towards miming the processes of analysis, evaluation, and human understanding. In fact, in the international scenario, in robotics, we find that the sensor system is at the service of artificial intelligence without solution of continuity and without elements that, in addition to sensations, can digitize perceptions, emotions, and affections. Furthermore, the same studies on affective computing appear in their own right to analyze specific issues and respond to particular needs. In this work, we instead present an integrated vision. The sensation captured by the sensor is transformed into perception, and then enriched with emotions and affections. Only after these steps, it will be possible to analyze the cognitive effect and the following decision, thanks to an artificial intelligence. The present work fills the gap between

sensing and artificial intelligence by modeling artificial human-inspired cognitive levels, so that we can define the “historical–cognitive enrichment of information” thanks to the stimulus–memories interaction.

In Figure 1, we want to introduce our info-structural model of cognition like an onion. Layers 1, 2, 3, and 4 depend on layers immediately outside, while Layers 5, 6, and 7 depend on the inside layers. The points of contact between two or more layers represent their retroactive dependencies.

1. Sensation
2. Perception
3. Emotion
4. Affection
5. Attention
6. Awareness
7. Consciousness

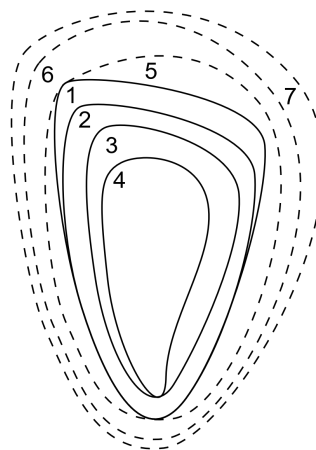


Figure 1. Intelligence structured as an onion whose layers are cognitive levels.

In the following sections, we present the model showing how cognitive state processing has been managed and how Smart Sensing has been modelled. The last part of the present study illustrates the experimental results obtained by processing visual stimuli inputs. In the discussion, cognitive levels related to the model are mentioned with the initial uppercase, while real human cognitive levels are mentioned with the initial lowercase. The contribution we want to convey is to suggest a general framework that acquires the five sensory signals to transform them into emotional and affective artificial cognition.

2. Related Works

The term “smart sensing” is commonly associated with applications regarding energy-efficient smart sensors, mostly for the Internet of Things [15,16]. In this paper, we want to adopt the same title to define the section of our framework that acquires sensory inputs to compute artificial instances of Sensation, Perception, Emotion, and Affection cognitive levels. This goal requires the definition of a cognition model which assumes an info-structural form.

The attempt of reproducing human perception, as well as emotion and affection, has been addressed in many ways, and it is difficult to illustrate each in a study that is not a review, but an original contribution in the direction of creating a basis to arrive at cognition starting from the sensation. We can, however, describe some of the most common approaches. In Reference [17], the concepts of sensation, perception, and cognition are taken into account separately, declining the second as active, i.e., sensory acquisition with environment adaptation capabilities, and passive, i.e., sensory acquisition without any feedback; in the case of an electronic tongue, modeling was performed according to a mapping of human perceptions to an artificial sensor system. The study takes into account levels of sensing, perception, and cognition; the immediate activity of our sensory system, the interpretation of sensory stimuli, and the acquisition, retrieval, and use of the information. The concept of attention is also taken into account for active perception as a function of the task being executed. The authors obtained two relevant results: a human-like electronic tongue, for evaluating food and water, and an artificial hand, for simulating the sense of touch. From a more biological point of view, tactile perception has also been emulated

through piezoelectric artificial synapses [18]. In Reference [19], texture features coinciding with human eye perception have been proposed, obtaining results that are comparable with the human visual sensory system. Visual perception has also been considered by [20], in which a cognitive system guides the attention to an object of interest, and it has been assumed simultaneous to goal strategies in [21], by validating a system through functional magnetic resonance imaging. Other interesting models of perception are those based on Bayesian frameworks [22] and artificial neural networks [23,24]. In general, except for [17], which explicitly considers a hierarchical model similar to the one presented in this study, all of the previous cited works regard particular sensory sources or concern the concept of perception by focusing on specific tasks such as feature extraction, object recognition, and localization.

In Reference [25], a hierarchical model based on multi-attribute group decision-making that combines personality, mood, and emotional states has been proposed. Here, personality is represented as a vector of characteristics, with mood as a state space in which the origin corresponds to the “neutral” and the affective model as the mapping between these states and the emotions. The hierarchy follows the order of personality, mood, emotion, and affective state. The results showed that, by creating a model based on group experts’ traits, it is possible to assist, or even to replace, the groups themselves in generating affective states for decision making. An example of an emotional framework based on Reinforcement Learning has been proposed in [26]. In this framework, agents learn cooperative behaviors. They receive rewards from the environment as a consequence of their actions, computing sensations through an internal environment composed of emotion appraisal and derivation models that generate intrinsic rewards useful for behavioral adaptation. Furthermore, affective interaction mechanisms have also been studied in [27]: social effects of emotions are classified primarily as emotions experienced but not communicated, emotions experienced and intentionally communicated, and emotions not experienced but intentionally communicated. Starting from a psychological background, they have realized a multi-agent system in which competitive and cooperative interactions have been obtained between agents with negative and positive social connections, respectively.

The purpose of this work was to generalize the attempt to reproduce the above cognitive levels, providing a framework that does not conceive an agent’s actions with respect to the environment, but only the stimuli acquisition through time-dependent polynomial functions, sliding window memories, and machine learning.

3. Proposed Model and Cognitive State Processing

As we can see in Table 1, it is possible to refine, through a computational fashion, in several cognitive levels, what happens when we make decisions, estimates, assessments, or recognize patterns. Specifically, this study refers to non-interacting agents—machines that perform no actions with respect to the external environment—and their hierarchy of cognitive levels.

In our model, the artificial agent intercepts environmental events by means of its Sensation level and processing sensory data, and sends its results to the next level. Each cognitive level receives and processes a set of data, producing a result, subsequently recorded in memory, that is forwarded to the next level of the hierarchy. Once the computations above have been completed, the agent obtains a tuple of results, whose components are kept in memory for a limited period of time.

A cognitive state c_n is defined as the tuple of results related to cognitive levels’ computations. Formally,

$$c_n = (r_{1,n}^{dec}, \dots, r_{i,n}^{dec}, \dots, r_{l,n}^{dec}) = (r_{1,n}^{dec}, \dots, r_{i,n}^{dec}, \dots, r_{7,n}^{dec}) \quad n \in \mathbb{N} \quad (1)$$

where $r_{i,n}^{dec}$ is the result, or instance, of the i -th cognitive level related to the acquisition, at the discrete time n , of an external event, and l is the number of cognitive levels in our structure, which is 7 in total.

Table 1. The info-structural model as an union of two macro-areas: Smart Sensing and Consciousness.

| Intelligence Activities for Non-Interacting Agents | | | | |
|--|-------------|--------------------|-------------------|-------------------------------|
| Sensing | | | | |
| Status | Implication | Memory | Functional Status | Generalized Functional Status |
| 1. Sensation | | | | |
| ↓ | ← | General Memory | Sensing | |
| 2. Perception | | | | |
| ↓ | ← | Emotive Memory | | Smart Sensing |
| 3. Emotion | | | | |
| ↓ | ← | Affective Memory | Sentiment | |
| 4. Affection | | | | |
| ↓ | | | | |
| Consciousness | | | | |
| 5. Attention | ⊃ | Self-Attention | | |
| ↓ | | | | |
| 6. Awareness | ⊃ | Self-Awareness | | Consciousness |
| ↓ | | | | |
| 7. Consciousness | ⊃ | Self-Consciousness | | |
| ↓ | | | | |
| Decision | | | | |

By “external event” we mean the sampling, at the instant n , with step T_{sens} , of the sensory input signals, since the sensory sphere of a robot is made of sensors. The above assumption seems to be reasonable since artificial agents, e.g., assume a visual capacity through cameras, which capture frames, i.e., samples of the visual reality. In this study, cognitive state acquisition and cognitive level processing is considered instantaneous and characterized by a non-limited capacity, pseudo-instantaneous behavior, and ideal parallelism/concurrency.

Human beings seem not to possess a memory capable of remembering, in time, every acquired cognitive instance. In fact, memory is often classified as short-term and long-term [28]. In order to simplify mathematically this characteristic, it is reasonable to ensure that cognitive instances would be volatile information. In fact, compared to an emotion, a sensation is less decisive for the subject’s decision; it is noticeable that an agent’s behavior is dependent on its emotional state, even when the sensory stimulus that elicited the given emotion is no longer captured [29]. The closer we get to the affection, the more cognitive levels seem to hold back cognitive data over time; emotions stimulate the activity of memory [30]. Thus, we consider cognitive instances $r_{i,n}^{dec}$ as information progressively decaying in agent’s memory.

The removal period of the i -th cognitive level instance is defined as the period T_i that elapses between the cognitive state acquisition and the elimination of $r_{i,n}^{dec}$ from the memory related to the i -th level. It must also respect the following bound:

$$T_1 < \dots < T_i < \dots < T_7, \quad (2)$$

where T_1, T_2, \dots, T_7 are the removal periods of the instances related to the seven cognitive levels. In this way, the agent's decisions depend more decisively on the activity of deeper cognitive levels. Indeed, human–environment interaction leads to decisions that depend more on an affection than on a sensation [31,32].

By way of example, being the cognitive state acquisition period T_C , the period that elapses between the first cognitive level's input acquisition and the last cognitive level's result generation, we show, considering l levels, some sequential steps concerning the cognitive state temporal processing:

1. an agent acquires a Sensation instance at time n ;
2. the agent processes the result of the i -th cognitive level at the time n , with $1 < i < l$;
3. the agent processes the result of the last cognitive level l at the instant $n + T_C$;
4. the agent removes the Sensation instance acquired at time n at the instant $n + T_S$, where $T_S = T_1$ is the removal period of the instance;
5. the agent removes the result recorded at time n of the i -th cognitive level at the instant $n + T_i$, with $1 < i < l$;
6. the agent removes the acquired result at time n of the last cognitive level l at the instant $n + T_l$.

Consequently, a new Sensation instance acquisition can take place at the instant $n + T_{sens}$.

4. Smart Sensing

Smart Sensing is modeled as a hierarchy of four cognitive levels, which are Sensation, Perception, Emotion, and Affection. Each level is thought as linked to the next one, providing as output a result, i.e., a cognitive instance. In order to support subsequent processing, each instance is periodically memorized into an evanescent window memory, disappearing after a predefined period of time.

To provide visual demonstrations on how the model works, we assume low dimensionality sensory inputs and cognitive instances. In a real use case, as we show in the last section of this study, cognitive instances can be highly dimensional and proportional to the deployed feature extraction. For example, to describe a sensory visual input, it could be possible to use a Convolutional Neural Network [33] to provide, in time, as many signals as needed to match the dimensionality of the features' vector. This aspect will be neglected in order to show some graphs and not to overly weigh down the notation.

As regards sensations and perceptions, the literature is generally discordant since researchers consider them as synonyms. In this study, these concepts are considered as not equivalent; specifically, the following hypothesis and definitions are adopted.

4.1. Sensation

Sensations allow our mind to understand ourselves and the world around us. Although they are essentially personal and subjective, it is impossible to measure them exactly, but it is possible to ask people to describe them. This first qualitative experiment makes possible to compare sensations and to note that, in some cases, they are caused by physical world specific changes, i.e., what is outside of us and what we perceive. Generally, every variation of the physical world is perceived by human beings in such a way that description of its variation is very similar.

Although the above premise seems obvious, it allows us to suppose that there are psychophysical relations between certain stimuli—physical variables—and some sensations—psychological variables—that tend to be predictable and independent with respect to an observer. To reach a further level of detail, it is possible to distinguish the concepts of “sensation” and “neuro-sensation”. The first declination is exactly the definition provided above and it is linked to sensory organs. On the other hand, when the stimulus, coming from sensory organs, reaches the central nervous system, more correctly, we must consider a “neuro-sensation”, or perception, a feeling that is enriched thanks to the memory of experiences. This explains, for example, the reason why people can derive a different perception from the same sensory stimuli. In our study, we will consider sensation and perception as two different cognitive levels.

In human beings, sensation is considered as the modification of our neurological system due to stimuli offered by the environment and captured by our sensory organs, whose channels are hearing, sight, smell, taste, and touch. In machines, a sensation can be considered as the sum of the contributions related to the sensors’ signals. For example, while for human beings the sense of sight is acquired through the activity of eyes, for a machine a similar result is acquired through a camera. The same goes for the hearing sense; what humans listen to through their ears can be computed by machines through their microphones.

While human beings’ sensation is a complex cognitive level to describe, for a machine its definition is simpler. For example, a machine’s visual sensation is made of pixels, each of which, independently from the color-orientation, can assume maximum and minimum values (corresponding, respectively, to white and to black, by considering the RGB format). The sensory input related to camera frames has huge dimensions since three signals are evaluated for each pixel in the scene, but their characteristics are practically the same. Similarly, machine auditory sensation is made of sampled waves’ intervals, each of which can assume a certain range of magnitude values on the decibel scale. These two examples of sensation can be considered as a computational approximation of those of human beings.

The additive sensory signal is then defined as the sum of the five sensory input signals. Formally,

$$S(n) = \sum_{i=1}^5 s_i(nT_{sens}) \quad n \in \mathbb{N}. \quad (3)$$

Sensory functions $s_i(t)$ are described by random processes with threshold, since the arrival of a sensation to the agent derives from stochastic occurrences. Likely, human beings perceive their conscious sensory experience as long as the stimuli overcome a certain threshold [34]; this phenomena is also noticeable when analyzing electrodermal activity (EDA) signals [35]. Therefore, we can define the threshold of the i -th sensory input $s_i(t)$ as follows:

$$H_{s_i} > 0. \quad (4)$$

The Sensation cognitive instance is defined as the vector of decaying sensory inputs. Formally,

$$r_{1,n}^{dec} = \mathbf{s}_n^{dec} = (dec(s_1(n), t), \dots, dec(s_i(n), t), \dots, dec(s_5(n), t)) \quad (5)$$

$$s_i(n) = s_i(nT_{sens}) \quad (6)$$

where $i = 1, \dots, 5$, $n \in \mathbb{N}$ and T_{sens} is the sampling period of the i -th sensory input.

Furthermore, Sensation memory m_s is defined as a tensor of the following type:

$$m_s = \begin{pmatrix} s_{n_f} \\ \vdots \\ s_{n_i} \\ \vdots \\ s_{n_h} \end{pmatrix} = \begin{pmatrix} s_1(n_f) & \dots & s_i(n_f) & \dots & s_5(n_f) \\ \vdots & & \vdots & & \vdots \\ s_1(n_i) & \dots & s_i(n_i) & \dots & s_5(n_i) \\ \vdots & & \vdots & & \vdots \\ s_1(n_h) & \dots & s_i(n_h) & \dots & s_5(n_h) \end{pmatrix} \quad (7)$$

with $n_f < n_i < n_h \in \mathbb{N}$ and whose terms $s_i(n)$ expire following the decay function:

$$dec(s_i(n), t) = A_s(s_i(n)e^{-\frac{t-n}{n}} - t_{0_{s_i}}) = A_s(s_i(nT_{sens})e^{-\frac{t-n}{n}} - t_{0_{s_i}}) \quad (8)$$

$$A_s = \frac{1}{1 - e^{-\frac{T_s}{n}}} \quad (9)$$

$$t_{0_{s_i}} = s_i(n)e^{-\frac{T_s}{n}} = s_i(nT_{sens})e^{-\frac{T_s}{n}} \quad (10)$$

$$s_i(n) < H_{s_i} \implies dec(s_i(n), t_b) = 0 \quad (11)$$

$$t_b < nT_{sens} \implies dec(s_i(n), t_b) = 0 \quad (12)$$

$$t_b > nT_{sens} + T_s \implies dec(s_i(n), t_b) = 0 \quad n \in \mathbb{N}. \quad (13)$$

with t_b as the generic continuous temporal instant.

Analyzing the total additive sensory signal, evaluated at discrete cognitive state acquisition time instants, it is possible to understand how the Sensation cognitive level processes sensory input signals together. The additive sensory signal tends to acquire the shape of the input with the highest magnitude.

In Figure 2, we highlight how Sensation computes the function $S(n)$ when processing both increasing and decreasing sensory signals, together with how its cognitive instances are managed in the general memory. Dashed lines, on the left picture, represent the time ranges during which the most recent sensory sample is acquired and the decay behavior of the cognitive instances is achieved. When the sixth state of the memory is acquired, the first sensory sample, recorded at the zero instant, has already been deleted from the general memory. When an instance is added to the memory, its intensity and its relevance has already started decaying.

Sensation has been modeled as a cognitive layer which takes, at every instant n , samples of the five sensory inputs and, according to their amplitude, computes the decay function $dec(s_i(n), t)$ to decrease their importance as a function of time. The values of the decay function are acquired by the Perception cognitive level, which keeps them in memory to combine. For each sensory dimension, the current sample is acquired, and the decay function relative to the previously acquired samples is computed.

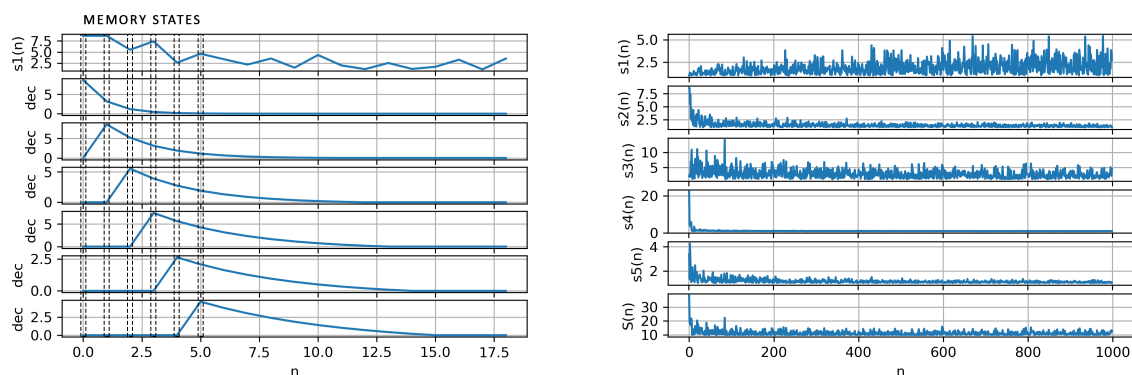


Figure 2. On the left, an acquired $s_1(n)$ is the sensory signal with six Sensation cognitive instances that decay in memory, with $T_S = 11$ s. On the right, $s_1(n)$ and the remaining four sensory signals are combined to form the total additive sensory signal $S(n)$.

4.2. Perception

The laws of perception are said to be autochthonous since they are considered as innate and not as a result of learning, even though an evolutionary progress in the elaboration of perceptions themselves is present. Since the first months of life, the newborn has been able to recognize colors and shapes—in particular, human figures—but only after the acquisition of the so-called “perceptive constancy”, i.e., the ability to connect forms or figures in which he recognizes similarities [36].

While in Europe, the Gestalt developed the phenomenological laws of perception, in the United States, the New Look of Perception paradigm took hold. The latter school, practically neglected by Gestalt, has become relevant to the personal and social values related to perceived objects. The forms are no longer considered innate and are anchored to the needs and purposes of individuals. Personal values and needs have become key elements in structuring perceptive processes, and significant objects and symbols are perceived as distorted and dissonant.

Under the above hypothesis, by leaving out more psychological aspects and orienting the discussion on a scientific-informative vision, we can introduce, from a systemic-functional point of view, the following definition.

Perception cognitive instance is defined as the following:

$$r_{2,n}^{dec} = P^{dec}(n, m_g^{dec}, \mathbf{s}_n, t) = dec(P(n, m_g^{dec}, \mathbf{s}_n), t) \quad (14)$$

$$m_g^{dec} = \begin{pmatrix} dec(P(n_f, m_{g_{n_f}}^{dec}, \mathbf{s}_{n_f}), t) \\ \vdots \\ dec(P(n_i, m_{g_{n_i}}^{dec}, \mathbf{s}_{n_i}), t) \\ \vdots \\ dec(P(n_h, m_{g_{n_h}}^{dec}, \mathbf{s}_{n_h}), t) \end{pmatrix} \quad n_f < n_i < n_h < n \in \mathbb{N} \quad (15)$$

$$dec(P(n, m_g^{dec}, \mathbf{s}_n), t) = A_p \left(\frac{\alpha^T \mathbf{s}_n}{t} + \frac{\beta m_g^{dec}}{t^2} - t_{0_{p_i}} \right) \quad (16)$$

$$\alpha = (\alpha_1(n), \dots, \alpha_i(n), \dots, \alpha_5(n)) = (\alpha_1(nT_{sens}), \dots, \alpha_i(nT_{sens}), \dots, \alpha_5(nT_{sens})) \quad (17)$$

$$A_p = \frac{\alpha^T \mathbf{s}_n + \beta m_g^{dec}}{\frac{\alpha^T \mathbf{s}_n}{n} + \frac{\beta m_g^{dec}}{n^2} - t_{0_{p_i}}} \quad (18)$$

$$t_{0_{p_i}} = \frac{\alpha^T \mathbf{s}_n}{n + T_p} + \frac{\beta m_g^{dec}}{(n + T_p)^2} \quad (19)$$

where $dec(P(n, m_g^{dec}, \mathbf{s}_n), t)$ is the Perception's decay from the memory m_g^{dec} , in time, following polynomial order. Weight vector β , related to memory, is a term that indicates, progressively, the relevance of window elements $m_{g,i}^{dec}$. If β_h is the weight related to the most recently added perceptive element in the window, and β_f is the weight related to the least recently added perceptive element in the window itself, it follows necessarily that $\beta_h > \beta_f, 0 < \beta_i < 1 \forall \beta_i \in \beta$. The quantity α is the weight vector linked to single Sensation instances, and β is the weight vector related to residents in memory Perception instances. Vector α is not constant, but variable, and it depends on time and on the results retrieved from the cognitive level of attention, since human beings' perception is affected by attention [37]. The terms A_p and $t_{0_{p_i}}$ have been introduced to obtain the desired behavior related to the decay of the instances, while the term related to the memory is divided by t^2 to achieve faster decay.

The elimination of a generic memory element $m_{g,i}^{dec}$ occurs at the instant $t = n_i + T_p$, where T_p is the removal period of the perceptual instance in memory.

Finally, we define the perceptive vector, useful to compute cognitive instances of the next layer of the model, as follows:

$$\mathbf{p}_n = (P_1^{dec}(n, m_g^{dec}, \mathbf{s}_n^{dec}, n), \dots, P_i^{dec}(n, m_g^{dec}, \mathbf{s}_n^{dec}, n), \dots, P_N^{dec}(n, m_g^{dec}, \mathbf{s}_n^{dec}, n)) \quad (20)$$

$$P_i^{dec}(n, m_g^{dec}, \mathbf{s}_n^{dec}, n) = \alpha_i(n T_{sens}) dec(s_i(n), n) + \sum_{j=f}^h \beta_j dec(P(n_j, m_{g_{n_j}}^{dec}, \mathbf{s}_{n_j}^{dec}, n), n). \quad (21)$$

The effect of the weight vector α , together with the contributions related to the designed memory model, can be observed analyzing the results of Perception. Figures 3 and 4 show an effective example of how Perception computes its results. By setting a short removal period T_p , this cognitive level does not increase the magnitude of its inputs, but models sensory signals according to the content of the general memory and to the state of the vector α .

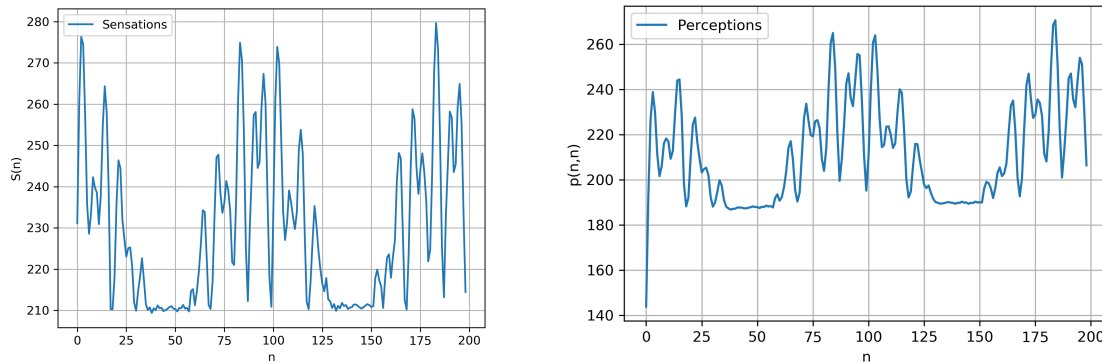


Figure 3. Agent total additive Sensation and Perception instances evaluated at discrete cognitive state acquisition time instants, with $T_s = 1$ s and $T_p = 3$ s. Here it is evident that, because of the general memory contribution, Sensation and Perception present different behaviors.

As the additive sensory signal, Perception also tends to acquire the shape of the input characterized by the highest magnitude. This result seems reasonable since human beings tend to perceive the most relevant received signals. For example, a loud sound increases the perception level with respect to the hearing sense [38], while a room with no light decreases the perception related to sight. Moreover, high frequency sensory signals play an important role in defining the output, since the human cognitive model regards them as highly emotional informative [39]. For example, the auditory sensory signal of a scream is characterized by a higher frequency, since human voice pitch is more acute.

Figure 3 shows a comparison between the total additive sensory signal and the Perception; on the right, perceptive peaks and the memory contribution between them are apparent. While flatter regions of $S(n)$ exhibit an oscillatory behavior, the Perception regions look smoother. In addition, as shown in Figure 5, Perception outputs always keep information about the sensory signals' shape, even when increasing the general memory capacity.

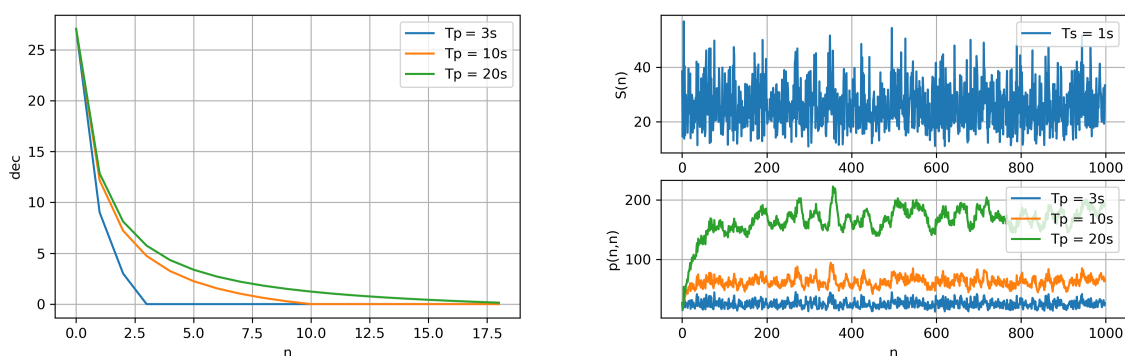


Figure 4. An example of how Perception cognitive instances polynomially decay in the general memory, following $T_p = 3s$, $T_p = 10s$, $T_p = 20s$ removal periods. Here, the weight vector is $\alpha = [0.9, 0.8, 1, 0.5, 0.6]$, and the Sensation instance removal period is $T_s = 1s$.

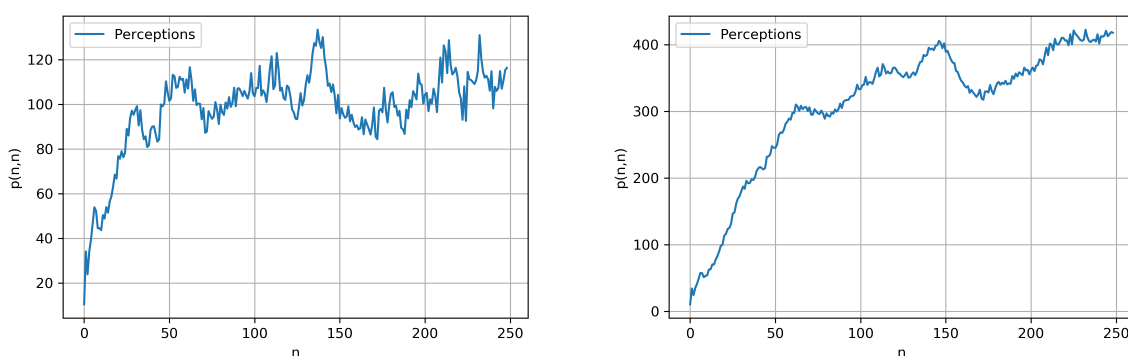


Figure 5. Agent Perception instances evaluated at discrete time instants, with removal periods $T_s = 1s$ and, respectively, $T_p = 15s$ and $T_p = 30s$. The Perception plot shows information related to Sensation, even when increasing its instance removal period.

Decays are computed as exponential functions characterized by the amplitude of acquired data. However, this statement is not necessarily true, since the sensory amplitude is modulated by the parameter α , which depends on the agent's level of attention related to the given input source. We think that this

model simplifies reasonably the way humans remember perceptions of external events through their senses, and takes effectively into account factors—memory and attention capacities—that are related to subjects' personal characteristics.

Perception takes the vector of Sensation cognitive instances \mathbf{s}_n computed in the previous level and, for each of its dimensions, applies time-varying weights stored in α to simulate the perception of a subject with respect to each sensory organ. For example, by considering α_1 as the weight related to the Perception of the sense of sight, its lowering towards zero indicates a visual deficit that can occur. This cognitive level applies the decay function $dec(P(n, m_g^{dec}, \mathbf{s}_n), t)$ to its inputs to decrease their importance as a function of time. All the instances of Perception acquired in the past are added to the current one to keep memory of the past. Finally, the values of the decay function are acquired by the Emotion cognitive level, which keeps them in memory to combine.

4.3. Emotion

In evolutionary or Darwinian terms, the main function of emotions is to make humans' reaction more effective during situations in which an immediate response is required for survival, i.e., a reaction with no necessary cognitive or conscious processing. According to Cannon-Bard, the emotional stimulus is firstly processed by subcortical centers of the brain, particularly by the amygdala, which receives information from thalamus posterior nuclei to induce an autonomic and neuroendocrine reaction. Emotions, though, cause many somatic modifications, e.g., heart rate change, increased or decreased sweating, respiratory rhythm acceleration, and muscle tension increases or relaxation.

Emotions also have a relational function, i.e., communicate and self-regulate our psychophysiological state. According to James-Lange's theory, emotion is a response to physiological variations. Humans experience many emotions with different physiological sensations and reactions. These theories have been criticized since people affected by spinal cord injuries still express emotions, as well as many similar physiological expressions. In some cases, especially due to strong emotions, a direct association, between physiological and emotional manifestations, still exists [40].

In order to build an effective model, it is necessary to take into account the classification of the most important emotions. One decisive contribution comes from the significant research conducted by Paul Ekman. He led thousands of experiments and acquired a high amount of data related to our topic of interest [41]. Therefore, we can classify emotions into anger, disgust, sadness, happiness, fear, surprise, and contempt [42].

The Emotion cognitive instance is defined as follows:

$$r_{3,n}^{dec} = E^{dec}(n, m_e^{dec}, \mathbf{p}_n, t) = dec(E(n, m_e^{dec}, \mathbf{p}_n), t) \quad (22)$$

$$m_e^{dec} = \begin{pmatrix} dec(E(n_f, m_{e_{n_f}}^{dec}, \mathbf{p}_{n_f}), t) \\ \vdots \\ dec(E(n_i, m_{e_{n_i}}^{dec}, \mathbf{p}_{n_i}), t) \\ \vdots \\ dec(E(n_h, m_{e_{n_h}}^{dec}, \mathbf{p}_{n_h}), t) \end{pmatrix} \quad n_f < n_i < n_h < n \in \mathbb{N} \quad (23)$$

$$dec(E(n, m_e^{dec}, \mathbf{p}_n), t) = A_e \left(\frac{\gamma^T \mathbf{p}_n}{t} + \frac{\delta m_e^{dec}}{t^2} - t_{0_{e_i}} \right) \quad (24)$$

$$\gamma = (\gamma_1(n), \dots, \gamma_i(n), \dots, \gamma_5(n)) = (\gamma_1(nT_{sens}), \dots, \gamma_i(nT_{sens}), \dots, \gamma_5(nT_{sens})) \quad (25)$$

$$\gamma_i(n) = \gamma_i(n-1)A^{dec}(n-1, m_{a_{n-1}}^{dec}, \mathbf{e}_{n-1}^{dec}, t) \quad (26)$$

$$A_e = \frac{\gamma^T \mathbf{p}_n + \delta m_e^{dec}}{\frac{\gamma^T \mathbf{p}_n}{n} + \frac{\delta m_e^{dec}}{n^2} - t_{0_{e_i}}} \quad (27)$$

$$t_{0_{e_i}} = \frac{\gamma^T \mathbf{p}_n}{n + T_E} + \frac{\delta m_e^{dec}}{(n + T_E)^2} \quad (28)$$

where $dec(E(n, m_e^{dec}, \mathbf{p}_n), t)$ is the Emotion's decay from the memory m_e^{dec} , in time, following polynomial order. Weight vector δ , related to memory, is a term that indicates, progressively, the relevance of window elements $m_{e,i}^{dec}$. If δ_h is the weight related to the most recently added emotional element in the window, and δ_f is the weight related to the less recently added emotional element in the window itself, it follows necessarily that $\delta_h > \delta_f$, $0 < \delta_i < 1 \forall \delta_i \in \delta$. The quantity γ is the weight vector linked to single Perception instances, and δ is the weight vector related to resident in memory Emotion instances. Vector γ is not constant, but variable, and it depends on time. The terms A_e and $t_{0_{e_i}}$ have been introduced to obtain the desired behavior related to the decay of the instances, while the term related to the memory is divided by t^2 to achieve faster decay.

The elimination of a generic memory element $m_{e,i}^{dec}$ occurs at the instant $t = n_i + T_E$, where T_E is the removal period of the emotional instance in memory.

Finally, we define the emotional vector, useful to compute cognitive instances of the next layer of the model, as follows:

$$\mathbf{e}_n = (E_1^{dec}(n, m_e^{dec}, \mathbf{p}_n^{dec}, n), \dots, E_i^{dec}(n, m_e^{dec}, \mathbf{p}_n^{dec}, n), \dots, E_N^{dec}(n, m_e^{dec}, \mathbf{p}_n^{dec}, n)) \quad (29)$$

$$E_i^{dec}(n, m_e^{dec}, \mathbf{p}_n^{dec}, n) = \gamma_i(nT_{sens})P_i^{dec}(n, m_g^{dec}, \mathbf{s}_n^{dec}, n) + \sum_{j=h}^f \beta_j dec(E(n_j, m_{e_{n_j}}^{dec}, \mathbf{p}_{n_j}^{dec}), n) \quad (30)$$

Results shown in Figures 6 and 7 represent sample plots of Emotion. It is evident that, while Perception presents evident fluctuations—mostly related to sensory signals' shape—Emotion radically attenuates this behavior by providing smoother functions. It is possible to discriminate emotional peaks and assign them to emotional classes. In addition, as illustrated in the figure, the increase in the emotional memory capacity, by increasing the removal period T_E , results in more relevant memory contribution and wider emotional peaks, but the differences with Perception are still evident.

We also observe that, for both Perception and Emotion, the increase in the memory capacity results in the increase in agent sensibility. This characteristic can be seen by comparing the plots in Figures 5 and 6: regarding the first plot, the function trend rapidly changes; regarding the second plot, the function stabilizes more persistently over time. This is an interesting result, since sensitive subjects are inclined to maintain perceptive and emotional states for a longer duration.

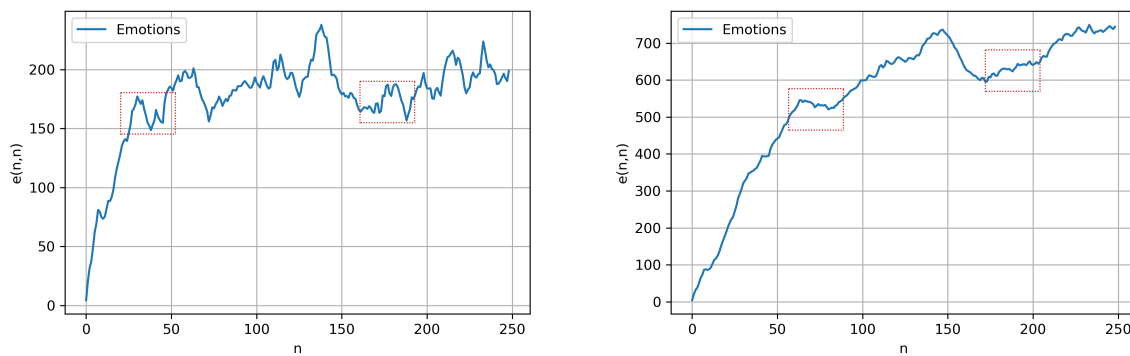


Figure 6. Agent Emotion instances evaluated at discrete time instants, with removal periods $T_S = 1$ s, $T_P = 3$ s and, respectively, 15 s and 30 s. The red squares highlight two intervals that show how agent emotional sensitivity changes when increasing/decreasing the removal period T_E . While in the first configuration, the agent rapidly changes Emotion levels, in the second one, it stabilizes its states, making them more durable and relevant. The plot on the right is similar to the Perception result, with $T_P = 30$ s, in Figure 5, but Emotion is smoother.

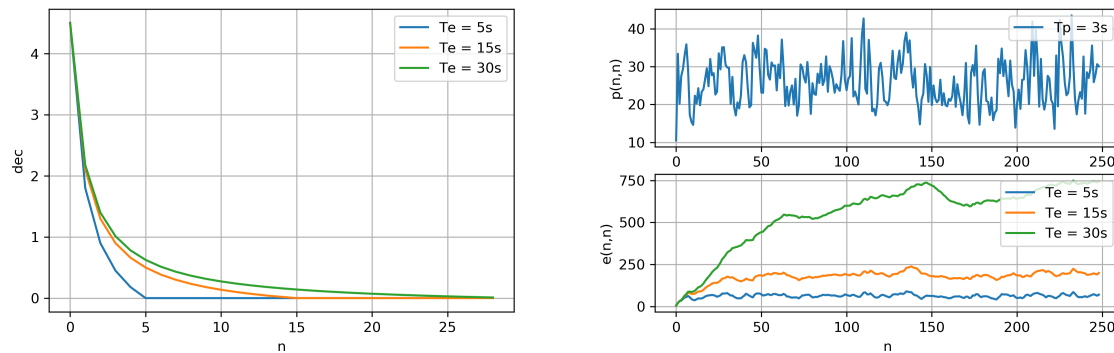


Figure 7. An example of how Emotion cognitive instances polynomially decay in the emotional memory, following $T_E = 5$ s, $T_E = 15$ s, $T_E = 30$ s removal periods. Here, the weight vector is $\gamma = [0.6, 0.7, 0.2, 1, 0.9]$, and Sensation and Perception instance removal periods are, respectively, $T_S = 1$ s and $T_P = 3$ s.

However, considering Emotion as a polynomial combination of Perception instances and of the emotional memory is reductive, since it is not possible to associate deterministically a cognitive instance to a given emotion. A stochastic approach is therefore required. In general, M being the family of stochastic models that perform the classification of emotions by means of a training set, we define the emotional class related to an Emotion cognitive instance as follows:

$$C'_{e,n} = M(\theta', \phi(r_{3,n}^{dec})) \quad C'_{e,n} \in E_n \quad (31)$$

$$E_n = (Ne_n, AR_n, DR_n, SaR_n, HR_n, CR_n, SuR_n, Fr_n) \quad (32)$$

with E_n the vector of probabilities, respectively, related to neutral, anger, disgust, sadness, happiness, contempt, surprise, and fear. $C'_{e,n}$ is the predicted emotional class at the instant n , while θ' is the model's set of parameters, and $\phi(r_{3,n}^{dec})$ is the basis function for the transformation of the cognitive instance at time n .

Training samples are therefore of the following form:

$$\langle E_n, r_{3,n}^{dec} \rangle \quad \forall n = 1, \dots, N_{train} \quad (33)$$

where $\langle E_n, r_{3,n}^{dec} \rangle$ indicates the association between Emotion cognitive instances $r_{3,n}^{dec}$ and each emotional class E_n , and N_{train} is the number of samples in the training set. We can obtain a vector of scores for every emotional class. This approach is clearly supervised since, to obtain a fully unsupervised emotional learning, it is necessary to classify Emotion cognitive instances through an autonomous activity of consciousness, since emotion discrimination requires at least a moral basis. However, in this discussion, we do not want to introduce such a complexity.

Emotion takes the vector of Perception cognitive instances \mathbf{p}_n computed in the previous level and, for each of its dimensions, applies time-varying weights stored in γ to simulate the emotion of a subject with respect to each sensory organ. For example, by considering γ_1 as the weight of the Emotion related to the sense of sight, its lowering towards zero indicates an emotional desensitization that can occur. This cognitive level applies the decay function $dec(E(n, m_e^{dec}, \mathbf{p}_n), t)$ to its inputs to decrease their importance as a function of time. All the instances of Emotion acquired in the past are added to the current one to keep memory of the past. Finally, each Emotion cognitive instance is associated with an emotional class and the values of the decay function are acquired by the Affection cognitive level, which keeps them in memory to combine. The weights in γ are also dependent on the decay function related to Affection cognitive instances acquired at the immediately previous instant.

4.4. Affection

Aristotle conceives affection as “páthos”—one of the ten categories of the substance; senses produce affections to impress sensory data on the spirit. The elements that cause sensitive and sentimental changes in the spirit, e.g., pleasure, pain, and desire, come from external objects; therefore, the affections coincide with the “passions” of the ethical sphere. The latter meaning is also found in Cicero, who adopts “affectiones” as a synonym of “perturbatio animi”, and in Augustine of Hippo, who uses the terms “perturbationes”, “affectus”, and “affectiones” as synonyms for “passiones”.

According to Plato, Cartesio, Spinoza, Leibniz, and Hegel, whereas good behavior is based on knowledge of truth, the affections are dangerous because they affect negatively cognition and moral attitudes. In the Aristotelian and Epicurean philosophies, the affections are valid in the cognitive field since sensory data are passively received by the subject and therefore they are always true, while anticipatory judgments are false. No man exists without passions; they need to be moderated instead of removed. Kant states it is essential that our spirit is “affected” by affections—otherwise, the cognitive activity of reasoning would be false—but if they are conceived as passions, their role is negative, i.e., cancers of practical reason.

For this work, with the aims of stimulating an advancement of ICT technologies by computationally deepening the above concepts, we provide the following definition.

The Affection cognitive instance is defined as follows:

$$r_{4,n}^{dec} = A^{dec}(n, m_a^{dec}, \mathbf{e}^{dec}_{n,n}, t) = dec(A(n, m_a^{dec}, \mathbf{e}_n), t) \quad (34)$$

$$m_a^{dec} = \begin{pmatrix} dec(A(n_f, m_{a_{n_f}}^{dec}, \mathbf{e}^{dec}_{n_f}), t) \\ \vdots \\ dec(A(n_i, m_{a_{n_i}}^{dec}, \mathbf{e}^{dec}_{n_i}), t) \\ \vdots \\ dec(A(n_h, m_{a_{n_h}}^{dec}, \mathbf{e}^{dec}_{n_h}), t) \end{pmatrix} \quad n_f < n_i < n_h < n \in \mathbb{N} \quad (35)$$

$$dec(A(n, m_a^{dec}, \mathbf{e}_n), t) = A_a \left(\frac{\mu^T \mathbf{e}_n}{t} + \frac{\lambda m_a^{dec}}{t^2} - t_{0_{a_i}} \right) \quad (36)$$

$$A_a = \frac{\mu^T \mathbf{e}_n + \lambda m_a^{dec}}{\frac{\mu^T \mathbf{e}_n}{n} + \frac{\lambda m_a^{dec}}{n^2} - t_{0_{a_i}}} \quad (37)$$

$$t_{0_{a_i}} = \frac{\mu^T \mathbf{e}_n}{n + T_A} + \frac{\lambda m_a^{dec}}{(n + T_A)^2} \quad (38)$$

where $dec(A(n, m_a, \mathbf{e}_n), t)$ is the Affection's decay from the memory m_a , in time, following polynomial order. Weight vector λ , related to memory, is a term that indicates, progressively, the relevance of window elements $m_{a,i}^{dec}$. If λ_h is the weight related to the most recently added affective element in the window, and λ_f is the weight related to the less recently added affective element in the window itself; it follows necessarily that $\lambda_h > \lambda_f$, $0 < \lambda_i < 1 \forall \lambda_i \in \delta$. The quantity μ is the weight vector linked to single Emotion instances, and λ is the weight vector related to resident in memory Affection instances. Vector μ is not constant, but variable, and it depends on time. The terms A_a and $t_{0_{a_i}}$ have been introduced to obtain the desired behavior related to the decay of the instances, while the term related to the memory is divided by t^2 to achieve faster decay.

The elimination of a generic memory element $m_{a,i}^{dec}$ occurs at the instant $t = n_i + T_A$, where T_A is the removal period of the emotional instance in memory.

Affection, as shown in Figures 8 and 9, heavily attenuates Emotion behavior with a sort of emotional peak grouping. This seems reasonable, since an affection can be considered as the synthesis of a certain set of emotions felt at a given time. Notice that the more the memory capacity increases, the more the Affection improves its emotional synthesis.

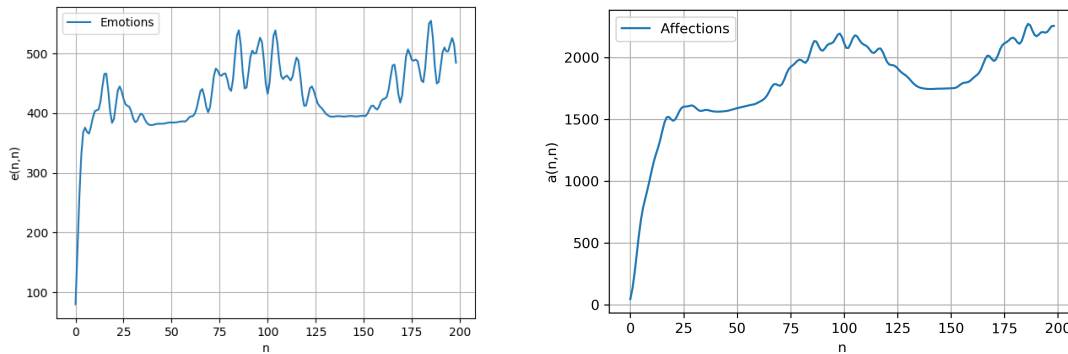


Figure 8. Agent Emotion and Affection instances evaluated at discrete time instants, with $T_E = 4$ s and $T_A = 10$ s.

Affection takes the vector of Emotion cognitive instances \mathbf{e}_n computed in the previous level and, for each of its dimensions, applies time-varying weights stored in λ to simulate the affection of a subject with respect to each sensory organ. For example, by considering γ_1 as the weight of the Affection related to the sense of sight, its lowering towards zero indicates an affection decrease that can occur. This cognitive level applies the decay function $dec(A(n, m_a^{dec}, \mathbf{e}_n), t)$ to its inputs to decrease their importance as a function of time. All the instances of Affection acquired in the past are added to the current one to retain memories from the past.

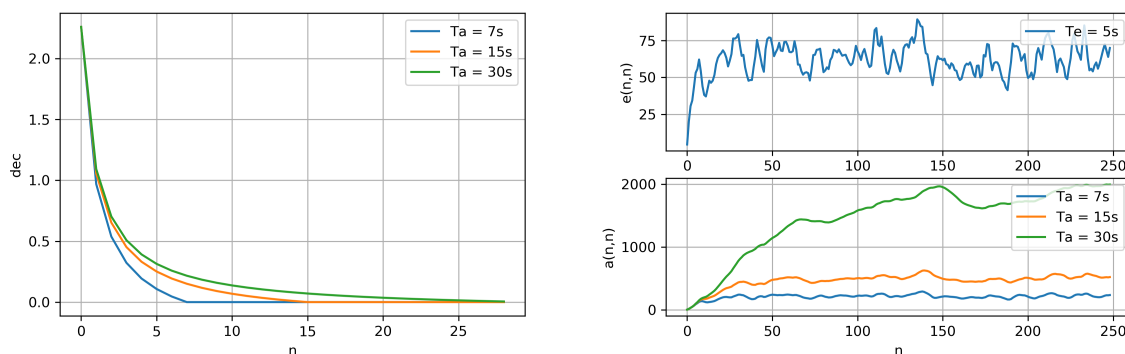


Figure 9. An example of how Affection cognitive instances polynomially decay in the affective memory, following $T_A = 7$ s, $T_A = 15$ s, and $T_A = 30$ s removal periods. Here, the weight vector is $\mu = [0.1, 0.9, 0.2, 0.8, 1]$, and Sensation, Perception, and Emotion cognitive instances removal periods are, respectively, $T_S = 1$ s, $T_P = 3$ s, and $T_E = 5$ s.

We have modeled affection as the decaying contributions of current emotions related to the sensory signals and the past affective history. This cognitive level is intended as a combination, a “grouping”, of different emotions related to an external entity—an “emotional synthesis”. Finally, we can conclude that, while Sensation and Perception can be classified as unique categories, what we call “Sensing”, Emotion, and Affection can be classified as another category, called “Sentiment”. Both categories are part of the macro-category called “Smart Sensing”.

5. Study Case and Framework Functionalities

A study case of the above model concerns the learning of a certain emotional behavior in relation to some sensory stimuli. Each cognitive instance can be classified with a combination of the seven main emotions illustrated in Section 4.3, obtaining an “oriented” emotional experience that depends on the agent’s perceptive and affective states. The main idea is to provide the agent with some stimuli and to impose the emotional reactions it must consequently activate. This requires building a dataset in which every configuration of the model, in terms of its parameters and sensory stimuli, is associated with a given emotional class. In this way, the agent, based on the provided experience and the Perception states recorded, will show emotions with respect to the stimuli he receives. Thus, every cognitive instance $r_{i,n}^{dec}$ is computed through the sum of sensory inputs with different dimensions; for example, the visual sensory input, as we will see in the following section, can have a dimensionality that depends on the adopted feature extraction. According to the past emotional experience obtained through a learning phase, the agent must provide emotions by adequately generalizing on sensory inputs; this characteristic permits the agent to infer, coherently with its memory state, an emotional behavior with respect to stimuli it has never seen before. For this reason, cognitive instances are classified according to the desired behavior by taking into account the parameters of each cognitive level. For example, in the case where the agent acquires a sensory input s_n^{dec} , with $\alpha_1 = 1$, since we want its reaction to be emotionally neutral, it is possible to determine, in the case of a subsequent input acquisition s_{n+1}^{dec} , with $\alpha_1 = 0.1$, that the next emotional reaction will still be neutral. This use case translates into the model the emotional reaction that a human being should present when its level of perception decreases. The Affection cognitive level provides an additional level of detail for the agent decision since, according to the content of its memory, it is possible to classify a set of emotions provided in the past with respect to one or more stimuli. However, in this

paper, we want to discuss the results achieved by focusing on the level of Emotion with respect to certain visual stimuli episodes.

6. Experiments with Visual Stimuli

The following experimentation is affected by the work in [7], in which spontaneous emotional activity was recognized and classified for a group of people by subjecting visual stimuli in order to create a database of facial expressions. The above study considered a real case of human emotional action, and we were inspired to try a similar approach by “replacing” one of those subjects with the model presented in this paper. A person receives a visual stimuli, e.g., an image, and accordingly shows an emotion; we endeavor to have an artificial agent that can acquire the same kind of visual stimuli and consequently to output emotions. Our purpose was not to accomplish a comparison between machines and individuals, but to instruct Smart Sensing to supply those emotions felt by human beings when they are affected by certain incentives. Thus, we present a use case in which sequences of images are transformed into Emotion cognitive instances in order to produce artificial emotional activity.

In Figure 10, we can find the learning architecture and a method to validate a learner coupled with the Emotion cognitive level. The present experiment is partitioned into two cores: (i) model evaluation, conducted by training and testing the learner with predefined episodes—a series of Emotion cognitive instances obtained by supplying pre-established sequences of visual stimuli; (ii) emotional activity, achieved by testing the formerly trained model on never-seen episodes—a series of Emotion cognitive instances obtained by supplying shuffled configurations of the visual stimuli served for the evaluation test. The former task is necessary to guarantee that the learner provides the desired emotions with respect to the established episodes; the latter is essential to inspect the emotions the learner produces when the memory m_e^{dec} presents states different from those involved in the evaluation.

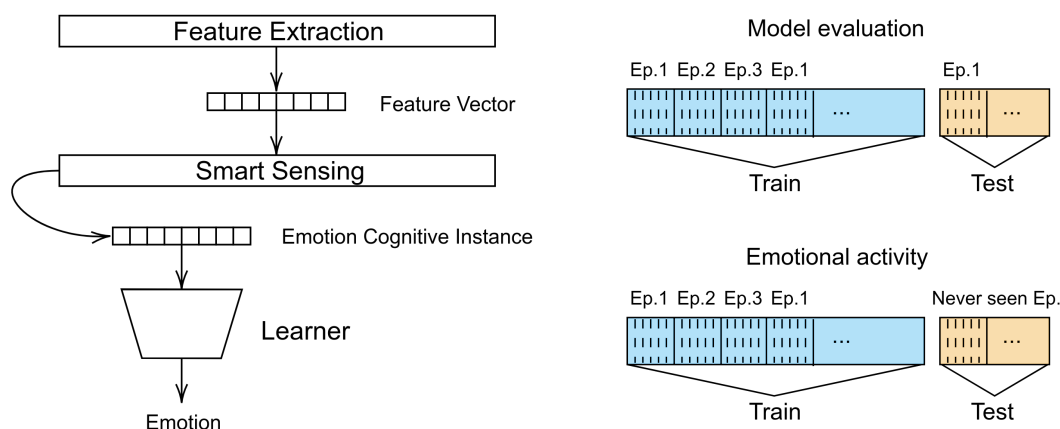


Figure 10. The emotional learning schema. Episodes represent given configurations of subsequent visual stimuli supplied in a predefined order. Model evaluation is performed by training and testing on episodes of three types, while emotional activity is assessed by testing the learner on the same test stimuli involved in the evaluation, but presented in a different order.

The images, reshaped to a 150×220 size, are labeled according to the approach described in (32) and are subject to the ImageNet [43] network, which provides features vectors of 12,288 components, of which an example of transformation into a cognitive instance is shown in Figure 11. After a first training phase, the learner is evaluated and re-trained subsequently to a feature selection based on importance weights. The evaluation is executed by training and testing on three types of episodes— $[0, 2, 3, 3, 5, 6]$, $[0, 7, 2, 1, 3, 4]$,

and $[0, 1, 7, 7, 4, 0]$ —whose stimuli are listed in Table 2. During the training phase, when a given emotion class becomes associated with an Emotion cognitive instance, the learner acquires the above association as a function of the current instance and the state of the Emotion memory. By way of example, when an Emotion cognitive instance associated with an injury stimulus becomes labeled with disgust, the agent will be trained to show disgust, with respect to the injury stimulus, by also keeping information, through the memory m_e^{dec} , about instances acquired previously. Thus, the emotional behavior depends on the order through which images are supplied to the agent.

Table 2. Stimuli collected for model evaluation and emotional activity, partially inspired by the study in [7]. These associations determine, taking into account Emotion memory m_e^{dec} effects, during the training phase, the emotional behavior the agent acquires at the time it captures a given sensory input.

| Class | Emotion | Stimuli | Relative Frequency |
|-------|-----------|---------------------------------------|--------------------|
| 1 | Happiness | Beautiful woman, own home | 0.11 |
| 2 | Anger | Murder of animal | 0.11 |
| 3 | Fear | War, man pointing weapon, terrorism | 0.18 |
| 4 | Surprise | Crazy sportsman | 0.11 |
| 5 | Contempt | Politician, parking car | 0.06 |
| 6 | Sadness | Someone's death or sick, car accident | 0.06 |
| 7 | Disgust | Injury, autopsy | 0.16 |
| 0 | Neutral | Landscape | 0.21 |

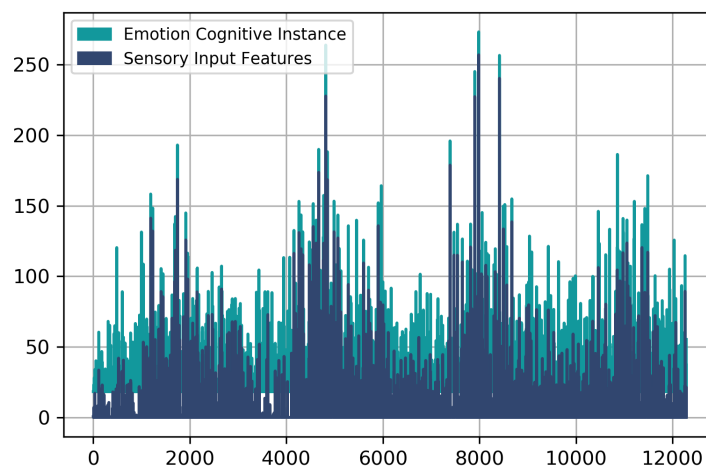


Figure 11. An example of how visual sensory input features are transformed into an Emotion cognitive instance. This plot describes the fifth image provided during the training phase of the learner; here, it can be seen how memory contributes in changing the content of the features vector when $T_S = 1$ s, $T_P = 1.5$ s, $T_E = 2$ s, $H_{S_i} = 0.001$, and $\alpha_1 = \beta_1 = \gamma_1 = 1$ and when there is an m_e^{dec} capacity at four locations.

Together with the accuracy, a “coherency” metric is also considered. It represents the ability of the model to distinguish between positive, negative, and neutral emotional stimuli. Positive stimuli are considered as associated with happiness and surprise, while neutral and negative ones, respectively, are associated with the neutral and the other remaining emotions. We decide to train the learner by means of ensemble models [44], in order to acquire more stable predictions for the small dataset we constructed—about 612 samples. In Table 3, we show the model evaluation performances obtained by using Random Forest [45,46] and XGBoost [47]; we reached better accuracy and coherency with gradient

boosting, of which a confusion matrix is shown in Tables 4 and 5. Contempt- and happiness-related stimuli are sometimes confused; the former is recognized as happiness, the latter as sadness. However, as can be seen in Table 5, confusion regarding positive and negative emotional stimuli turns out to be suitable for the experiments. The results of the emotional activity on never-seen episodes involve accuracy and coherency reduction due to the effect of memory. In fact, without the contribution of m_e^{dec} , evaluation and emotional activity tests provide the same results.

Table 3. Evaluation tests, with $T_S = 1$ s, $T_P = 1.5$ s, $T_E = 2$ s, $H_{s_i} = 0.001$, $\alpha_1 = \beta_1 = \gamma_1 = 1$, and an m_e^{dec} capacity at four locations.

| Learner with Random Forest | | | | | | | |
|----------------------------|---------------|-----------|----------|----------|------------|----------|-----------|
| Samples | Training size | Test size | Episodes | Features | Estimators | Accuracy | Coherency |
| 348 | 296 | 53 | 58 | 3654 | 250 | 0.67 | 0.75 |
| 438 | 373 | 66 | 73 | 980 | 200 | 0.74 | 0.84 |
| 612 | 520 | 92 | 102 | 3378 | 500 | 0.78 | 0.86 |
| Learner with XGBoost | | | | | | | |
| Samples | Training size | Test size | Episodes | Features | Estimators | Accuracy | Coherency |
| 348 | 296 | 53 | 58 | 3654 | 100 | 0.77 | 0.86 |
| 438 | 373 | 66 | 73 | 980 | 100 | 0.80 | 0.82 |
| 612 | 520 | 92 | 102 | 3378 | 200 | 0.85 | 0.92 |

Table 4. Confusion matrix of the accuracy related to the model trained with XGBoost over 612 samples, with $T_S = 1$ s, $T_P = 1.5$ s, $T_E = 2$ s, $H_{s_i} = 0.001$, $\alpha_1 = \beta_1 = \gamma_1 = 1$, and an m_e^{dec} capacity at four locations.

| Predicted | | | | | | | | |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Neutral | Happiness | Anger | Fear | Surprise | Contempt | Sadness | Disgust |
| Neutral | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Happiness | 0.05 | 0.77 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 |
| Anger | 0.00 | 0.00 | 0.90 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fear | 0.05 | 0.00 | 0.09 | 0.89 | 0.00 | 0.00 | 0.00 | 0.00 |
| Surprise | 0.00 | 0.00 | 0.00 | 0.05 | 0.88 | 0.00 | 0.00 | 0.00 |
| Contempt | 0.05 | 0.11 | 0.00 | 0.00 | 0.00 | 0.57 | 0.00 | 0.08 |
| Sadness | 0.05 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.57 | 0.08 |
| Disgust | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.91 |

Table 5. Confusion matrix of the coherency related to the best model trained with XGBoost over 612 samples, with $T_S = 1$ s, $T_P = 1.5$ s, $T_E = 2$ s, $H_{s_i} = 0.001$, $\alpha_1 = \beta_1 = \gamma_1 = 1$, and an m_e^{dec} capacity at four locations.

| Predicted | | | |
|-----------|-------------|-------------|-------------|
| | Negative | Neutral | Positive |
| Negative | 0.92 | 0.16 | 0.05 |
| Neutral | 0.00 | 1.00 | 0.00 |
| Positive | 0.03 | 0.05 | 0.83 |

In Table 6, we show the emotional activity tests performed over episodes different from those used for training; the model was tested over a test set composed of the same samples used for the evaluation, but shuffled randomly. The contribution of m_e^{dec} is noticeable; it determines a lowering of accuracy and

coherency, causing a completely different emotional activity, compared to the emotions predicted in the evaluation test. Even if the agent shows different emotions towards the same kind of stimuli, the relative frequency of predicted emotions is approximately equal to the relative frequency obtained for the evaluation. For example, as can be seen in Table 6, even when the agent, during the emotional activity, has predicted different emotions with respect to the same visual stimuli provided for the evaluation test, it always shows approximately 23% of Fear. The obtained results show the dependence of the emotional activity on m_e^{dec} and on the order through which visual stimuli are provided to the model. The agent outputs emotions according to its past history and with an approximately constant relative frequency.

Table 6. Emotional activity related to the best model trained with XGBoost over 612 samples, with $T_S = 1$ s, $T_P = 1.5$ s, $T_E = 2$ s, $H_{s_i} = 0.001$, $\alpha_1 = \beta_1 = \gamma_1 = 1$, and an m_e^{dec} capacity at four locations. The system achieves a first emotional generalization responsiveness of 94%.

| | Relative Frequency | | | | | | | | |
|-------------------------|--------------------|-----------|-------|------|----------|----------|---------|---------|-----------|
| | Neutral | Happiness | Anger | Fear | Surprise | Contempt | Sadness | Disgust | Acc/Coh |
| Evaluation | 0.24 | 0.09 | 0.12 | 0.23 | 0.09 | 0.04 | 0.05 | 0.14 | 0.85/0.92 |
| Emotional activity 1 | 0.25 | 0.10 | 0.12 | 0.23 | 0.07 | 0.05 | 0.02 | 0.15 | 0.80/0.89 |
| Emotional activity 2 | 0.22 | 0.10 | 0.09 | 0.23 | 0.09 | 0.04 | 0.06 | 0.16 | 0.80/0.92 |
| Emotional activity 3 | 0.25 | 0.11 | 0.11 | 0.22 | 0.08 | 0.04 | 0.02 | 0.17 | 0.82/0.91 |
| Mean emotional activity | 0.24 | 0.10 | 0.11 | 0.23 | 0.08 | 0.04 | 0.03 | 0.16 | 0.81/0.91 |

The results of the emotional activity as a function of γ_1 , shown in Figure 12, suggest that, when the Emotion visual weight γ_1 tends toward 0.1, accuracy and coherency over never-seen episodes assume the same score achieved for the evaluation test (85% of accuracy and 92% of coherency).

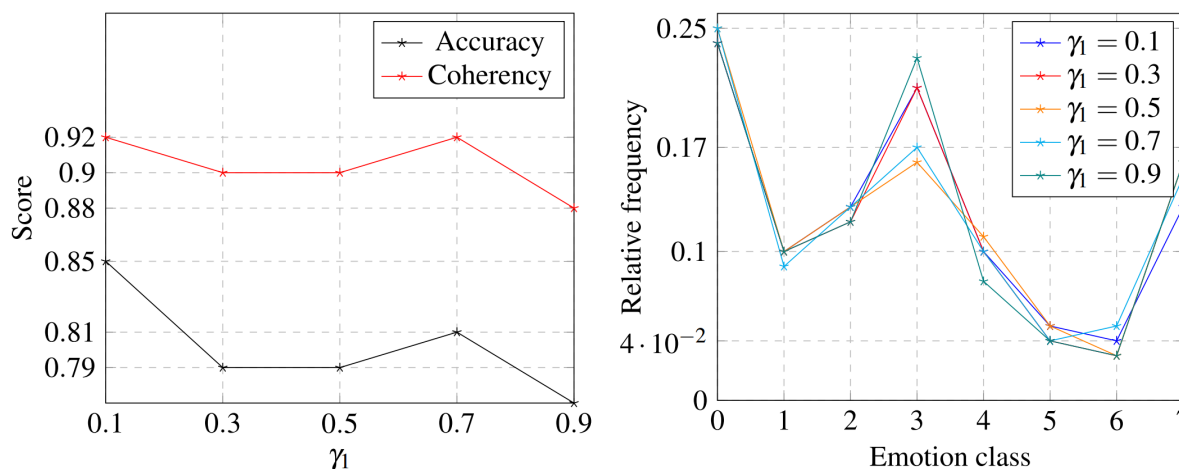


Figure 12. Emotional activity as a function of γ_1 related to the best model trained with XGBoost over 612 samples, with $T_S = 1$ s, $T_P = 1.5$ s, $T_E = 2$ s, $H_{s_i} = 0.001$, $\alpha_1 = \beta_1 = 1$, and an m_e^{dec} capacity at four locations.

When the importance of Emotion cognitive instances decreases, the agent tends to ascribe less relevance to current memory m_e^{dec} states for the emotional activity, predicting emotional classes according to the past state of the memory related to the training phase. During the emotional activity, when γ_1 decreases, the agent takes into account the emotional history to which it has ascribed more importance in the past.

7. Discussion

Results show that the agent acquires the desired emotional experience with suitable accuracy and without the use of a neural network. The emotional behavior is consistent by virtue of the constant relative frequency of emotions; the content of cognitive memories affects the emotional output and, as a consequence of the tests conducted on episodes different from those learned in the training phase, the agent provides different emotional behavior with respect to different sequences of stimuli. The variation in the emotional behavior is represented by the lowering of accuracy with respect to the initial model evaluation. We have also shown that the agent, by lowering the importance of the current emotion, behaves as if the new sequence of stimuli was identical to the one learned in the past. Therefore, when the agent does not ascribe importance to its emotions, it starts assuming a behavior that does not depend on the current stimuli but on the most emotionally relevant past ones.

A comparison between our model and the related works presented in Section 2 reveal differences. The study conducted in [17] introduced a threshold concept that is similar to what we defined in (4), but that associates it with perception rather than with sensation; we have assumed this process as prior to the cognitive acquisition of the Perception cognitive level. In Reference [17], the concept of artificial perception is also conceived as subjected to the assessment of a level of cognition that provides meaning to sensory stimuli. On the contrary, in our info-structural model, we associate the level of cognition with that of consciousness, which we assume takes into account also emotional and affective processing related to the perceived stimuli. We believe, in fact, that human cognition also depends on the emotional memory associated with those stimuli. Furthermore, our perception representation is closer to the concept of passive perception, since our model does not take into account the agent's actions towards the environment. Other studies on perception [18–24] have very few similarities with the present study, since they focus on particular characteristics by addressing problems that we have overlooked due to the generalization aims of the present study. In fact, our intention was to model the general functioning of perception in relation to the cognitive levels of Emotion and Affection. The study in [25] models a personality characteristics vector as a function of variables related to neuroticism, extraversion, openness, conscientiousness, and agreeableness, which are instantiated by answering personality tests. In our model, the tracts of personality are determined by the emotional training occurring during the model evaluation theorized in Section 4.3 and described experimentally in Section 6. In Smart Sensing, an agent's personality is determined dynamically by the associations between the emotional classes and the perceptions of stimuli with the addition of related emotional memory content. This is, in fact, a pre-defined orientation through which we define the way the agent emotionally reacts to the environment. The approach in [26] regards an agent's behavioral adaptation as a function of the rewards received from the environment. In our framework, the emotional behavior does not depend on the actions the agent performs, since non-interacting emotional activity does not seem to depend on rewards. In fact, as shown in Section 6, our agent, like a human being, outputs negative emotions with respect to negative stimuli that, like the vision of someone sick, could not depend on its actions. Even though it is necessary to model an emotional mechanism based on the interaction with the environment, we think that it is more suitable to first build a cognitive model based on human behaviors that do not depend on rewards. Furthermore, we consider the adaptation to be associated with the conception the agent has of good and evil, our understanding of which will deepen in continuation of our studies regarding consciousness, in which we will investigate the level of human morality that does not depend on the actions performed in the past.

The experiment described in Section 6 can be placed in the scenario related to the research on the so-called “artificial emotion” [48,49]—emotions “felt” by a machine. In this field, remarkable studies include those of [50], in which robots are provided with facial expressions based on the interaction with a human partner, and [51], which provides a general framework for designing emotions in autonomous

agents. With the present study, our contribution allows an empirical and info-structural representation of artificial cognitive instances in terms of vectors. Machine emotions become “recognizable” inside an artificial agent and contribute to determining a form of emotional activity that depends on the upper cognitive levels—Sensation and Perception—and on cognitive memories. The present model acts regardless of agent goals and actions towards the environment, providing the agent with emotions even when it acts just as a “viewer”. Smart Sensing could be used as an empirical hypothesis for inquiring the way humans process information perceptively and emotionally, as well as verifying future emotional activities of subjects downstream to a specific history of stimuli. Smart Sensing represents the starting point of a new approach for implementing an artificial consciousness, which takes into account Sensation, Perception, Emotion, and Affection cognitive levels as a function of time. In addition to demonstrating that the model functions in reproducing artificial emotional activity, providing good results with a limited amount of samples, this study provides a framework, never presented before, enforced by psychological and behavioral literature, which encourages the development of artificial intelligence systems through the observation of human sensations, perceptions, emotions, and affections. We argue that, at the current stage of research in artificial intelligence, it is no longer suitable to design cognitive systems that neglect the highlighted cognitive levels and the inter-functional relationships between them. As we have seen, emotion depends on the way the stimuli are perceived through sensory sources and on the level of affection related to the stimuli themselves; this is a statement revealed to be true for human beings, thus testable by everyone, and should be true also for artificial “minds”. The research community can contribute to the present research line by performing experiments that also take other sensory channels into account, e.g., auditory or tactile, and by expanding Smart Sensing with further cognitive levels. From an applicative point of view, through our implementation choices, it is possible to neglect the use of neural networks, which, by definition, need large amounts of samples, when developing emotional sensor systems, e.g., agents with affective capacity. Future developments could consist in trying to turn the learner used to output emotional classes into a formal model that does not include any form of supervision, which is a challenging task. The solution to this last issue may involve the dependence of the Emotion cognitive level on a mechanism of consciousness capable of distinguishing emotionally positive stimuli from negative and neutral ones. Another future extension of this work is the support of cognitive instance classification with a facial expression recognition model, a smart way of providing the agent with empathic functionalities. The present work lays the foundations for the design of our idea of artificial consciousness, which will be addressed in a subsequent paper. We will present an info-structural model of cognition based on attention, awareness, and consciousness, which, as we will see, intrinsically depends on Smart Sensing.

We intend to underline the novelty of this study from a methodological point of view. It opens a unitary perspective regarding the interaction between all the cognitive levels reproducible in an artificial agent. This is a great challenge for the research community, since it also favors the concurrence of the human sciences and the overall branches of knowledge, which, in a logic of integration, can offer a renewed technological–scientific and humanistic path in the present moment of research, elaboration, and applications. The present work intends to outline a research hypothesis and illustrate horizons according to an overall vision of the subject-man/artificial agent relationship.

8. Conclusions

In this work, we presented an initial reflection on the human–machine relationship and traced a possible methodology through which to design a computational model able to provide emotions to an artificial agent. We described our hierarchical idea of a cognitive model regarding intelligence activities for non-interacting agents, showing Smart Sensing layers’ mathematical translation. We showed how cognitive levels of Sensation, Perception, Emotion, and Affection were modeled, as well as their characteristics and

interactions. In doing so, we described how the model is supported by evanescent window memories. We also exposed the experimental results obtained by running the model with simple test input signals, analyzing numerical outcomes, and highlighting the differences between cognitive levels' outputs in terms of their parameters and characteristics. In the last section, taking inspiration from a real human case, we illustrated a demonstration of the model functioning on the task of producing artificial emotions, obtaining results that are suitable in terms of emotional generalization and reasonable from a behavioral point of view. Following an analysis of human aptitudes, we begin to deepen our understanding of the cognitive levels of attention, awareness, and consciousness and of how Smart Sensing potentials improve when an artificial consciousness comes into play.

Author Contributions: Conceptualization, G.I., I.F., R.E.L., and F.T.; methodology, G.I., I.F., R.E.L., and F.T.; software, G.I., I.F., R.E.L., and F.T.; data curation R.E.L.; writing—original draft preparation G.I., I.F., R.E.L., and F.T.; supervision, G.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Poria, S.; Cambria, E.; Bajpai, R.; Hussain, A. A review of Affective Computing: From unimodal analysis to multimodal fusion. *Inf. Fusion* **2017**, *37*, 98–125. [CrossRef]
2. Lee, Y.; Park, C.; Choi, H. Word-level emotion embedding based on semi-supervised learning for emotional classification in dialogue. In Proceedings of the IEEE International Conference on Big Data and Smart Computing, Kyoto, Japan, 27 February–2 March 2019; pp. 1–4.
3. Ismail, S.; Ahmad, A.; Jawad, M.A. Human identity verification via automated analysis of fingerprint system features. *Int. J. Innov. Comput. Inf. Control.* **2019**, *15*, 2183–2196.
4. Maria, E.; Matthias, L.; Sten, H. Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55.
5. Ko, Y.; Hong, I.; Kim, Y. Construction of a database of emotional speech using emotion sounds from movies and dramas. *Int. Conf. Inf. Commun.* **2017**, *343*, 266–267.
6. Huang, Z. An investigation of emotion changes from speech. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Xi'an, China, 21–24 September 2015; pp. 733–736.
7. Zhalehpour, S.; Onder, O.; Akhtar, Z.; Erdem, C.E. BAUM-1: A spontaneous audio-visual face database of affective and mental states. *IEEE Trans. Affect. Comput.* **2017**, *8*, 300–313. [CrossRef]
8. Oh, Y.H.; See, J.; Ngo, A.C.L.; Phan, R.C.W.; Baskaran, V.M. A survey of automatic facial micro-expression analysis: Databases, methods and challenges. *Front. Psychol.* **2018**, *9*, 1128. [CrossRef] [PubMed]
9. Jagini, N.P.; Rao, R.R. Exploring emotion specific features for emotion recognition system using PCA approach. *Int. Conf. Intelligent Computing Control. Syst.* **2017**, *9*, 58–62.
10. Jang, E. H.; Park, B.J.; Kim, C.H.; Sohn, J.H. Emotion classification based on physiological signals induced by negative emotions: Discrimination of negative emotions by machine learning algorithm. In Proceedings of the 9th IEEE International Conference on Networking, Sensing and Control, Beijing, China, 11–14 April 2012; pp. 283–288.
11. Liu, G.D.; Li, Y.C.; Zhang, W.; Zhang, L. A brief review of artificial intelligence applications and algorithms for psychiatric disorders. *Engineering* **2020**, *6*, 462–467. [CrossRef]
12. Scheutz, M. Philosophical Issues about Computation. Available online: <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470018860.s00209> (accessed on 25 July 2020).
13. Shirowzan, S.; Sepasgozar, S.; Samad, M.E. Digital Twin and CyberGIS for Improving Connectivity and Measuring the Impact of Infrastructure Construction Planning in Smart Cities. *Int. J. Geo-Inf.* **2020**, *9*, 240. [CrossRef]
14. Lee, R.S.T. Smart City. In *Artificial Intelligence in Daily Life*; Springer: Singapore, 2020; pp. 321–345.

15. Zhang, S.; Li, S.; Cheng, S.; Ma, J.; Chang, H. Research on smart sensing RFID tags under flexible substrates in printed electronics. In Proceedings of the 2015 16th International Conference on Electronic Packaging Technology (ICEPT), Changsha, China, 11–14 August 2015; pp. 1006–1009.
16. Alahi, M.E.E.; Ishak, N.P.; Mukhopadhyay, S.C.; Burkitt, L. An internet-of-things enabled smart sensing system for nitrate monitoring. *IEEE Internet Things J.* **2018**, *5*, 4409–4417. [[CrossRef](#)]
17. Robertsson, L.; Iliev, B.; Palm, R.H.; Wide, P. Perception modeling for human-like artificial sensor systems. *Int. J. Hum. Comput. Stud.* **2007**, *65*, 446–459. [[CrossRef](#)]
18. Kumar, M.; Singh, R.; Kang, H.; Kim, S.; Seo, H. An artificial piezotronic synapse for tactile perception. *Nano Energy* **2020**, *73*, 104756. [[CrossRef](#)]
19. Jian, M.; Dong, J.; Gao, D.; Liang, Z. New texture features based on wavelet transform coinciding with human visual perception. In Proceedings of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007), Qingdao, China, 30 July–1 August 2007; pp. 369–373.
20. Yu, Y.; Mann, G.K.I.; Gosine, R.G. A goal-directed visual perception system using object-based top-down attention. *IEEE Trans. Auton. Ment. Dev.* **2011**, *4*, 87–103. [[CrossRef](#)]
21. Wang, R. Combined goal recursion strategy and visual perception strategy on human problem solving and cognitive system simulation. In Proceedings of the 2010 Third International Symposium on Information Science and Engineering, Shanghai, China, 24–26 December 2010; pp. 247–250.
22. Ferreira, J.; Castelo-Branco, M.; Dias, J. A Bayesian framework for active artificial perception. *IEEE Trans. Cybern.* **2013**, *43*, 699–711. [[CrossRef](#)] [[PubMed](#)]
23. Leitner, J.; Harding, S.; Frank, M.; Förster, A.; Schmidhuber, J. Artificial neural networks for spatial perception: Towards visual object localisation in humanoid robots. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–7.
24. Shan, G.; Wang, T.; Li, X.; Fang, Y.; Zhang, Y. A Deep Learning-based Visual Perception Approach for Mobile Robots. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 825–829.
25. Su, C.; Gao, Y.; Jiang, B.; Li, H.; Srikanta, P. An affective cognition based approach to multi-attribute group decision making. *J. Intell. Fuzzy Syst.* **2018**, *35*, 11–33.
26. Yu, C.; Zhang, M.; Ren, F.; Tan, G. Emotional multiagent reinforcement learning in spatial social dilemmas. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 3083–3096. [[CrossRef](#)]
27. Pudane, M. Affective multi-agent system for simulating mechanisms of social effects of emotions. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), San Antonio, TX, USA, 23–26 October 2017; pp. 129–134.
28. Izquierdo, I.; Medina, J.H.; Vianna, M.R.M.; Izquierdo, L.A.; Barros, D.M. Separate mechanisms for short- and long-term memory. *Behav. Brain Res. Sci.* **1999**, *103*, 1–11. [[CrossRef](#)]
29. Guojiang, W.; Xiaoxiao, W.; Kechang, F. Behavior decision model of intelligent agent based on artificial emotion. In Proceedings of the 2nd International Conference on Advanced Computer Control, Shenyang, China, 27–29 March 2010; pp. 185–189.
30. Phelps, E.A. Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.* **2004**, *14*, 198–202. [[CrossRef](#)]
31. Dolan, R.J. Emotion, cognition and behavior. *Science* **2002**, *298*, 1191–1194. [[CrossRef](#)]
32. Thierry, G.; Roberts, M.V. Event-related potential study of attention capture by affective sounds. *Neuroreport* **2007**, *18*, 245–248. [[CrossRef](#)]
33. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [[CrossRef](#)] [[PubMed](#)]
34. Libet, B.; Alberts, W.W.; Wright, E.J.; Feinstein, B. Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science* **1967**, *158*, 1597–1600. [[CrossRef](#)] [[PubMed](#)]
35. Kim, K.H.; Bang, S.W.; Kim, S.R. Emotion recognition system using short-term monitoring of physiological signals. *Med. Biol. Eng. Comput.* **2004**, *42*, 419–427. [[CrossRef](#)]

36. Godino, A.; Canestrari, R. La Psicologia Scientifica: Nuovo Trattato di Psicologia Generale. Available online: <https://www.torrossa.com/it/resources/an/2250598> (accessed on 10 July 2020).
37. Barutchu, A.; Toohey, S.; Shivdasani, M.N.; Fifer, J.M.; Crewther, S.G.; Grayden, D.B.; Paolini, A.G. Multisensory perception and attention in school-age children. *J. Exp. Child Psychol.* **2019**, *180*, 141–155. [CrossRef]
38. Rohl, M.; Uppenkamp, S. Neural coding of sound intensity and loudness in the human auditory system. *J. Assoc. Res. Otolaryngol.* **2012**, *13*, 369–379. [CrossRef]
39. Rodero, E. Intonation and emotion: Influence of pitch levels and contour type on creating emotions. *J. Voice* **2011**, *25*, 25–34. [CrossRef]
40. Bear, M.; Connors, B.; Paradiso, M.A. *Neuroscience: Exploring the Brain*; Jones & Bartlett Learning: Burlington, MA, USA, 2020.
41. Ekman, P.; Davidson, R.J. *The Nature of Emotion: Fundamental Questions*; Oxford University Press: Oxford, UK, 1994.
42. Ekman, P.; Keltner, D. Universal Facial Expressions of Emotion. Available online: <https://www.paulekman.com> (accessed on 12 October 2020).
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
44. Bauer, E.; Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting and variants. *Mach. Learn.* **1999**, *36*, 105–139. [CrossRef]
45. Pal, M. Random forest classifier for remote sensing classification: Bagging, boosting and variants. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]
46. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2005**, *25*, 197–227. [CrossRef]
47. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. *arXiv* **2016**, arXiv:1603.02754.
48. Wang, Z.; Xie, L.; Lu, T. Research progress of artificial psychology and artificial emotion in China. *CAAI Trans. Intell. Technol.* **2016**, *1*, 355–365. [CrossRef]
49. Kowalczyk, Z.; Czubenko, M. Computational approaches to modeling artificial emotional overview of the proposed solutions. *CAAI Front. Robot. AI* **2016**, *3*, 21.
50. Hara, F. Artificial emotion of face robot through learning in communicative interactions with human. In Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication, Okayama, Japan, 22–22 September 2004; pp. 7–15.
51. Rosales, J.; Rodríguez, L.; Ramos, F. A general theoretical framework for the design of artificial emotion systems in Autonomous Agents. *Cogn. Syst. Res.* **2019**, *58*, 7–15. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).