Citation for published version


DOI

http://doi.org/10.1109/ROMAN.2015.7333699

Link to record in KAR

http://kar.kent.ac.uk/56882/

Document Version

Author's Accepted Manuscript

Copyright & reuse
Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research
The version in the Kent Academic Repository may differ from the final published version. Users are advised to check http://kar.kent.ac.uk for the status of the paper. Users should always cite the published version of record.

Enquiries
For any further enquiries regarding the licence status of this document, please contact: researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at http://kar.kent.ac.uk/contact.html
An Architecture for Emotional and Context-Aware Associative Learning for Robot Companions

Caroline Rizzi Raymundo¹, Colin G. Johnson² and Patricia A. Vargas³

Abstract—This work proposes a theoretical architectural model based on the brain’s fear learning system with the purpose of generating artificial fear conditioning at both stimuli and context abstraction levels in robot companions. The proposed architecture is inspired by the different brain regions involved in fear learning, here divided into four modules that work in an integrated and parallel manner: the sensory system, the amygdala system, the hippocampal system and the working memory. Each of these modules is based on a different approach and performs a different task in the process of learning and memorizing environmental cues to predict the occurrence of unpleasant situations. The main contribution of the model proposed here is the integration of fear learning and context awareness in order to fuse emotional and contextual artificial memories. The purpose is to provide robots with more believable social responses, leading to more natural interactions between humans and robots.

I. INTRODUCTION

Emotional memory is essential to any social being. The same argument applies to robotic companions that are intended to establish long-term relationships with human users. A common issue in the human-robot interaction (HRI) field is the rapid lost of interest from users due to the robot’s lack of intelligent and adaptive response. Users get frustrated and lose motivation over time as companions continue to perform pre-defined and repetitive behaviors.

This problem must be tackled in order to produce a more engaging and natural interaction between robots and humans. From the perspective of social intelligent robots, Dautenhahn [1] argues that the better computational agents can meet our human cognitive and social needs, the more familiar and natural they are, and the more effectively they can be used as tools.

We believe that emotional memory is vital for long-term robot companions to be capable of learning and adapting to its dynamic environment. Neuroscientific findings indicate that emotions are strongly related with cognitive processing and play an essential role in functions considered crucial for intelligent behavior, such as fast decision-making, learning, perception and creativity [2], [3]. In addition, a robot capable

to express emotional responses may generate a richer social-interaction experience for humans [4].

For a social robot to effectively interact with humans, as important as being “emotionally appraised” is to be able to understand its context, as well as how it influences its own wellbeing. As humans, we expect others to be able to identify environmental factors that can represent unpleasantness or danger to themselves and act accordingly (eg., avoiding it). Therefore, predicting and reacting to these situations can highly increase the believability of a robot companion’s social behavior [5].

Human contextual memory and processing has always been source of inspiration for researchers in artificial intelligence, and their efforts for designing memory models has strongly contributed to improve our understanding of human cognition [6] [7] [8] [9] [10] [11]. In this direction, a strong contribution comes from the work of Kolodner [12], which represents one of the first comprehensive works in knowledge-based systems, allowing intelligent systems to remember situations as cases and extract reasoning rules from them. In addition, results of recent research emphasize the role of episodic memory on cognitive robots [13] and simulated agents [14], [15] through the learning and remembering of temporally sequenced episodes/events and their significance.

In the area of artificial fear learning, we highlight the model proposed by Morén and Balkenius [16]. Their model has been widely used in a large range of engineering and robotic applications where the system is required to adapt to environmental changes [17], [18], [19], [20]. However, contrary to our proposal, their model does not incorporate the concept of situation, thus creating “emotional” associations only at stimulus level.

This paper proposes an architecture based on the brain’s fear learning system. The aim is to generate artificial emotional conditioning, more specifically fear conditioning, at both stimuli and context levels. Through stimuli fear-conditioning, the robot shall form implicit “memories of fear” and react to stimuli that may predict the occurrence of unpleasant stimuli. The main contribution of our work is the introduction of the situation concept and its integration with implicit fear memories in other to form even more complex memories that represent a portion of reality as a whole.

This paper is organized as follows: the neuroscientific background for this work is briefly discussed in Section II, followed by the presentation of the proposed architecture in Section III. Finally, we discuss conclusions and future work in Section IV.

¹The first author is financially supported by CAPES, a Brazilian research-support agency (process number 0648/13-2).
²Caroline Rizzi Raymundo is PhD researcher at the School of Computing of the University of Kent, Canterbury, UK. c.rizzi-raymundo@kent.ac.uk
³Dr. Colin G. Jonson is Professor at the School Computing of the University of Kent, Canterbury, UK. c.g.johnson@kent.ac.uk
³Dr. Patricia A. Vargas is the Director of the Robotics Laboratory and Associate Professor at the School of Mathematical and Computer Sciences of the Heriot-Watt University, Edinburgh, UK. p.a.vargas@hw.ac.uk
II. THE EMOTIONAL BRAIN

In this section we discuss the neuroscientific background behind our theoretical model, detailing four brain systems essential to the situation-aware fear learning process in animals.

A. Sensory System

The sensory thalamus and sensory cortex, two important brain areas of the sensory system, are responsible for collecting and preprocessing environmental stimuli, which are later relayed to other brain regions for further processing. The main purpose of this preprocessing step is to inform higher order systems of the brain about the state of the environment in an abstraction level that these systems can understand.

For instance, the amygdala, which is the brain area believed to be the home for fear learning [3], does not receive sensory stimuli directly from the environment. Instead, it receives input by way of the thalamic and cortical pathways, which together compose the “low and high roads” to the amygdala respectively [3] (Fig. 1).

Because it is shorter, the thalamic pathway can provide the amygdala with low latency information about environmental stimuli, but without the aid of cortical processing. Therefore, although faster, information reaching the amygdala by means of the thalamic pathway provides a crude representation of the sensed stimulus. On the other hand, information projected through the thalamic-cortical pathway takes longer to reach the amygdala, but provides a higher level representation of the sensed world. As noted by LeDoux [3], “The Beatles and Rolling Stones (...) will sound the same to the amygdala by way of the thalamic projections but quite different by way of the cortical projections”.

The direct thalamic pathway suggests that fear responses can occur without the involvement of higher processing systems of the brain such as the cortex, thus bypassing brain regions involved in reasoning and consciousness. This direct pathway may be responsible for mediating emotional responses that we cannot consciously control and allows us to start reacting to dangerous situations even before we fully understand what is happening. The indirect cortical pathway, in turn, is responsible for overriding the signals from the thalamic pathway with more accurate information about the environment, so that we can consciously reason over it and adequate our response to the real danger.

B. Amygdala System

Fear learning is believed to take place mainly at the amygdala [3], [21], [16], and is mediated by plasticity mechanisms similar to those involved in Pavlov’s classical conditioning [22]. In other words, repeatedly pairing a neutral stimulus with an aversive unconditioned stimulus (which naturally produces fear or anxiety) triggers plasticity in the amygdala and generates an association between the neutral stimulus and the feeling of fear. The neutral stimulus then becomes a conditioned stimulus, which now carries emotional value.

In addition, associative learning in the amygdala is believed to be aided by inhibitory projections from the orbitofrontal cortex [16]. The amygdala’s role is to associate neutral stimuli paired with aversive stimuli in order to learn to predict that aversive stimuli. When the amygdala predicts an aversive stimulus, it outputs an adequate emotional response (i.e., fear) that will prepare the individual to deal with eminent danger. However, if for some reason the aversive stimulus becomes absent, then the learned association becomes invalid and the resulting emotional response becomes inadequate. The orbitofrontal cortex aids the amygdala by detecting the omission of expected noxious stimuli and inhibiting the amygdala’s response in proportion to the mismatch.

C. Hippocampal System

Context can be defined as a collection of stimuli and the relationship among them [3]. While sensory stimulus may refer to light levels, noise patterns, etc., context represents the part of reality surrounding us and comprises all detected stimuli and their meaning together during a given period of time.

The hippocampus is believed to be the main brain region involved in processing and representing contextual information [3]. It is where we begin to leave the purely perceptual reasoning about the world and enter the conceptual domain of the brain. In the hippocampus, the highly processed sensory information projected by the sensory cortex is mixed together in order to form a representation of the world that is no longer just visual, auditory or olfactory, but that includes all of these at once.

In addition, the hippocampus also receives emotional feedback from the amygdala through the release of the adrenaline hormone. When the amygdala detects an aversive stimulus, it activates a variety of bodily systems, including the autonomic nervous system, which in turn leads to the release of adrenaline. It then influences systems that are active at the same time. As consequence, explicit memories being formed in the hippocampus at the same time adrenaline is released get strengthened. Fig. 2 shows a rough representation of the interactions between the sensory, amygdala and hippocampal systems.

![Fig. 1. The low and high roads to the amygdala [3].](image-url)
D. Working Memory

In the fear learning process, the amygdala and hippocampal systems work in parallel to form different types of memories: the implicit (or unconscious) and explicit (or declarative) memories respectively. After a traumatic episode, both memories are reactivated if exposed to stimuli that were present during the trauma [3].

The hippocampus will retrieve contextual memories related to that trauma, such as who you were with, where you were and what you were doing at the moment of the trauma. In addition, it will remember you, as a cold fact, that the situation was unpleasant. At the same time, the amygdala will activate bodily responses characteristic of fear expression (e.g., tense up muscles, increase heart rate, etc.). Because these two systems are activated in parallel by the same stimuli, implicit and explicit memories seem to be part of a unified memory system.

The place where these two kind of memories meet is called working memory, and is also where immediate conscious experience is created. In the working memory, explicit memory of stored knowledge from past experiences and implicit emotional memory are fused in consciousness, so that newly formed explicit memories can be given emotional imprint.

III. Theoretical Model

Based on the neuroscientific background presented in Section II, we propose a theoretical architectural model of an artificial fear learning system capable of generating emotional conditioning at both stimulus and contextual abstraction levels. Note that with this model we do not attempt to capture all the real neural circuits involved in the brain’s fear learning system, which are far more complex and have not yet been completely understood by neuroscience. The proposed model seeks to capture the aspects of the fear learning system that are relevant for providing robots with more natural social responses.

Fig. 3 depicts the complete model of the proposed artificial fear learning system, showing how the four main modules of our architecture are interconnected. The sensory system, composed of the sensory thalamus and sensory cortex, preprocess environmental stimuli detected by the system (e.g., by means of sensors’ input or direct user input), which is then relayed to the amygdala and hippocampal systems. The amygdala system, composed of the amygdala and orbitofrontal cortex, is responsible for predicting and associating environmental stimuli to eminent aversive situations. It also provides emotional feedback to the hippocampal system, which in turn generates complex contextual representations of the environment based on the highly processed sensory information projected by the cortex. Finally, implicit memories from the amygdala system and explicit memories from the hippocampal system meet in the working memory, where contextual information is associated with emotional information to produce emotional responses.

Because the four modules have different goals and perform different tasks inside the architecture, each of them is based on a different approach that best suits their individual needs. In the sequel we enter in the details of each module, describing their respective approaches.

A. Artificial Sensory and Amygdala Systems

In our architecture, the modules representing the sensory and the amygdala systems are both based on the model proposed by Morén and Balkenius [16]. Fig. 4 depicts their fear learning model, which is based on the joint work of artificial neural networks (ANNs) representing the four brain regions discussed in Sections II-A and II-B, which are: sensory thalamus, sensory cortex, amygdala and orbitofrontal cortex.

In their model, the sensory thalamus and sensory cortex are modeled in a very simple way, having as only function superficially processing sensory input and relaying it to further
components of the model. The work of Morén and Balkenius [16] follows LeDoux’s “low and high roads” theory discussed in Section II-A. Therefore, the thalamus sends short-latency crude sensory information to the amygdala, while the cortex projects sensory information with higher level of abstraction to both the amygdala and the orbitofrontal cortex.

The main part of their model is divided between the amygdala and the orbitofrontal cortex. The amygdala is responsible for learning to predict sensory input based on associations made with reinforcing signals, which in our case are aversive stimuli. The orbitofrontal cortex, in turn, compares the expected reinforcing signal with the actual reinforcement. It then inhibits the amygdala’s response in proportion to the mismatch between expected and actual reinforcement, so to avoid inadequate emotional responses.

For example, suppose an autonomous vacuum cleaner capable to predict eminent aversive situations based on the amygdala model described in this section. Suppose that, during its cleaning, it gets tangled in the wires of a computer placed in a specific location of the room. After getting tangled a few times, through the amygdala system the robot (vacuum cleaner) associates that location in the room with getting tangled and starts to avoid it. But if the computer is moved to another room, the reinforcing signal becomes absent and there is no longer a reason for fearing that location of the room. It is the job of the orbitofrontal cortex to detect the absence of the aversive stimulus and to inhibit the amygdala’s emotional response. If the computer is moved back to the same location in the future, inhibition from the orbitofrontal cortex should decrease, so that emotional response from the amygdala becomes strong again.

**B. Artificial Hippocampal System**

The main job of the hippocampus is to extract the meaning of detected stimuli in order to represent as a whole the current state of affairs of interest. For instance, suppose a work environment that conducts weekly meetings every Tuesday 2PM. Suppose also that this work place benefits from a smart autonomous vacuum cleaner, which is capable to understand and adapt to its surroundings based on environmental feedback, including human interaction feedback.

Consider, in this scenario, that the robot (vacuum cleaner) has detected at certain moment the following set of environmental cues: week day: Tuesday; time: 2PM; location: meeting room; light levels: high; noise levels: medium to high; door status: closed. Alone, each of these stimuli has no valuable meaning, but together they can lead the robot to infer the context, which may be “there is a ongoing meeting in the meeting room”.

Systems that can perceive and understand the context of their surroundings are known as context-aware or even situation-aware systems. More specifically, situations represent an extension of the context concept in the sense that context-aware systems focus on characterizing the environment, whereas situation-aware systems seek to “determine when relevant entities are in a particular state so they (systems) can take action” [23]. For the goals of our fear learning architecture, we are particularly interested in situation-awareness approaches.

We have based the design of the hippocampus module on the tools available by SCENE [24], [25], which is a robust situation-management platform that builds on the JBoss Drools rule engine and its CEP (Complex Event Processing) platform [26] to provide support for developing rule-based situation-aware systems. Rule-based languages are based on the human cognitive process of conscious decision-making, which is guided by the rules and facts learned during an individual’s life [27]. This makes of rule-based techniques, with the aid of SCENE’s situation-management support, the suitable tool for simulating the hippocampal functions in the brain.

1) **Rules-Based Languages - Overview**: Rule-based programming languages consist of a set of rules (if-then statements) that can be repeatedly applied to a set of facts, which in turn represent immutable entities of the real world. The entity “meeting-room door” is an example of instance of the fact “door”, which may have as attribute its current status, i.e., “open” or “closed”. An example of rule using the “door” fact would be “if the meeting-room door has status equals closed then stop cleaning”.

The inference engine, also known as rule engine, is the component that evaluates facts against rules’ patterns through a process known as pattern matching. When one or more facts satisfy a rule’s condition (the if part), the inference engine executes the actions defined in the rule’s then part.

Some rule-based systems are able to handle events, which are records of significant changes in the domain’s state at a given point in time [26]. Some examples of events are “light has been turned on”, “it is 2PM now”, “I (robot) have entered the meeting room”, etc. A system can be notified of an event through sensors’ input or through the detection of a set of events, case in which the event is said to be a complex event. Because events have intrinsic temporal properties, they can be compared with each other by means of temporal
operations. For example, we may desire to create an event “meeting time” when the event “it is 7PM” happens after (which is the temporal operation) the event “it is Tuesday” has happened, considering that the event “it is Wednesday” has not yet occurred.

While events represent punctual changes in the state of affairs, such as “I have entered the meeting room”, situations represent changes in the state of affairs that have duration, such as “I am in the meeting room”. The life cycle of this particular situation starts when the event “I have entered the meeting room” is detected and ends when the event “I have left the meeting room” is detected. When a situation’s life cycle ends, it is considered to be a past situation. Temporal operations can also be performed among situations, as well as between events and situations. For example, we may have the situation “I am in the meeting room” during (which is the temporal operation) the period in which the situation “lights are on” is happening.

2) Situation Awareness Applied to the Hippocampus Module: The highly processed stimuli information projected by the sensory cortex is basically the system’s interpretation of the features of detected environmental events. Therefore, any new information sent by the sensory cortex to the hippocampus should be stored as the attribute of a new event detected by the inference engine. For the architecture to be domain independent, generic sensory information coming from the cortex should continue generic when stored as event in the memory of the CEP platform.

Adrenaline information coming from the amygdala follows similar principles, being a type of event as well. However, instead of representing a set of environmental features, as does sensory stimuli, the adrenaline actually depicts the intensity level of appraised fear. Therefore, it should be stored in a way it can be compared in magnitude, such as an integer value depicting the adrenaline’s current level.

Inside the fear learning architecture, situations are only valuable when they can represent any kind of danger or unpleasantness to the system. Fear caused by aversive stimuli, in turn, always increases adrenaline, whose level depends on how dangerous or unpleasant that aversive stimulus is. Thus, in our architecture, aversive situations start to exist when levels of the adrenaline signal (i.e. when the attribute “level” of an instance of the event “adrenaline”) sent by the amygdala rises above a given threshold, and ends when adrenaline levels return to normal. Such situation should comprise all cortical events detected during its life cycle, since they all contribute to capturing the state of the environment during that period of fear.

However, it is important to note that the purpose of our fear learning architecture is to provide the system with the ability to predict eminent aversive situations based on past experiences. Thus, even more important than capturing the state of the environment during the aversive situation is to capture its state before the aversive situation, so that in the future the system can use this information to predict these situations before their actual occurrence. To capture the state of affairs before the occurrence of the aversive situation, we propose the creation of a complex event at the moment adrenaline levels rise above the stipulated threshold that would comprise all events created in the last $x$ time units before the adrenaline signal rose (i.e., before the situation started), where $x$ should be defined by the system’s designer.

Fig. 5 summarizes the internal architecture of the module depicting the hippocampal system. Highly processed stimuli is projected by the sensory cortex to the hippocampus, at the same time adrenaline signal is sent from the amygdala to the hippocampus. Inside the hippocampus, the CEP platform processes stimuli information and transforms it into generic events. Finally, situations and complex events caring information about the state of affairs before and during aversive situations are generated depending on adrenaline levels.

In summary, at one side the hippocampus receives individual stimuli information as input and at the other side it outputs a collection of events composed of stimuli information and their temporal properties and relationships.

C. Artificial Working Memory

As explained in section II-D, the working memory is the place where implicit memory formed in the amygdala and explicit memory formed in the hippocampus are fused to create conscious emotional memories. This module represents the last piece of the puzzle for generating associative learning regarding situational and emotional information. In the working memory, an associative mechanism similar to classical fear conditioning [22] takes place. However, instead of conditioning the perception of individual stimuli to the feeling of fear, as done by the amygdala, we are actually conditioning the perception of context to the feeling of fear. By doing so, we allow the system to recover “memories of fear” by detecting environmental features present during or previous to the “feeling of fear” in past experiences.

For instance, consider the meeting example given in Section III-B. Suppose that, when the robot interrupts an ongoing meeting for the first time, a person reproves it for starting cleaning. Suppose also that this negative feedback is internally interpreted by the robot’s amygdala system as an unpleasant stimulus. At the same time, the hippocampal system collects all environmental features detected before and during this embarrassing situation. When information

![Fig. 5. Hippocampus internal architecture.](image-url)
coming from the amygdala and the hippocampus arrives at the same time in the working memory, a learning mechanism is triggered, which associates all collected situational patterns projected by the hippocampus with the emotional response projected by the amygdala. In the next time the robot’s hippocampal system detects the same environmental features (e.g., doors of meeting room closed at Tuesday 2PM, lights of meeting room on, etc.), the working memory will recall the emotional response associated with it. As consequence, the robot should hesitated entering the room in an attempt to avoid an emotionally unpleasant situation, thus preventing interrupting an ongoing meeting once again.

IV. CONCLUSIONS

We have proposed an architecture based on the brain’s fear learning system that aims at generating artificial emotional conditioning at both stimulus and contextual abstraction levels. The architecture is composed of four main modules: the sensory system, the amygdala system, the hippocampal system and the working memory.

For future work, we intend to implement the proposed architecture using the approaches suggested in this paper. More specifically, we intend to use classical ANNs to implement the model proposed by Morén and Balkenius [16], which comprises the sensory and amygdala systems. For implementing the hippocampal system, we intend to use SCENE [24] in conjunction with the Drools’ rule engine and CEP platform [26].

To implement the described associative learning mechanism of the working memory, we intend to use our previous work, ASP (Artificial Synaptic Plasticity) [28]. ASP is a mechanism that extends the classical feedforward ANN for simulating associative learning, which is conducted through a conditioning-like procedure. Learning takes place at runtime with ASP, so the robot can learn and create new associations while exploring the environment, which is an essential requirement in the proposed fear learning architecture. ASP was originally developed for handling association between coming from the amygdala and the hippocampus arrives at the same time in the working memory, a learning mechanism is triggered, which associates all collected situational patterns projected by the hippocampus with the emotional response projected by the amygdala. In the next time the robot’s hippocampal system detects the same environmental features (e.g., doors of meeting room closed at Tuesday 2PM, lights of meeting room on, etc.), the working memory will recall the emotional response associated with it. As consequence, the robot should hesitated entering the room in an attempt to avoid an emotionally unpleasant situation, thus preventing interrupting an ongoing meeting once again.

IV. CONCLUSIONS

We have proposed an architecture based on the brain’s fear learning system that aims at generating artificial emotional conditioning at both stimulus and contextual abstraction levels. The architecture is composed of four main modules: the sensory system, the amygdala system, the hippocampal system and the working memory.

For future work, we intend to implement the proposed architecture using the approaches suggested in this paper. More specifically, we intend to use classical ANNs to implement the model proposed by Morén and Balkenius [16], which comprises the sensory and amygdala systems. For implementing the hippocampal system, we intend to use SCENE [24] in conjunction with the Drools’ rule engine and CEP platform [26].

To implement the described associative learning mechanism of the working memory, we intend to use our previous work, ASP (Artificial Synaptic Plasticity) [28]. ASP is a mechanism that extends the classical feedforward ANN for simulating associative learning, which is conducted through a conditioning-like procedure. Learning takes place at runtime with ASP, so the robot can learn and create new associations while exploring the environment, which is an essential requirement in the proposed fear learning architecture. ASP was originally developed for handling association between...

REFERENCES