# The Role of Emotions in Autonomous Social Agents

**Luc Caspar**

Supervisor: Professor Roger K. Moore

Department of Computer Science
The University of Sheffield

PhD Thesis submitted for the degree of
*Doctor of Philosophy*

Speech and Hearing Research Group February 2020

# Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than $75,000$ words excluding appendices, bibliography, footnotes, tables, and equations and has fewer than 150 figures.

Luc Caspar
February 2020

# Abstract

The ‘*holy grail*’ of both AI and cognitive science is human-level intelligence. Whereas AI relies on computer algorithms to simulate human abilities, cognitive scientists investigate the brain to understand the underlying mechanisms. For most of human history, emotions were thought to be nothing more than disturbances for cognition. Therefore, they were usually ostracized from research on intelligence. As a result, cognitive architectures only partially include emotions in their design, if at all. Recently, though, it was discovered that emotions and cognition are in fact inter-dependent systems. Consequently, before being able to fully replicate human-level intelligence, it is necessary to understand the concept of emotions and its many roles within the brain. In this thesis, working around the lack of definition for emotion, I show that emotions inform the brain as to the nature of a given situation and guide the decision-making process, to increase the survival potential of virtual agents. In particular **ProtoEmo**, an architecture replicating the circuits found at the base of the forebrain, is shown to have the ability to detect stimuli relevant to the survival of virtual agents. Hence, it outperforms other emotional agents in terms of survival capabilities, which are measured by the size of the remaining population at the end of a resource foraging task. **PrimEmo**, the architecture born from the integration of **ProtoEmo** with standard models of the reward and decision-making systems in the brain, displays survival capabilities similar to the advantage actor-critic algorithm. **PrimEmo** also shows promises for supporting primitive emotions characterized by their level of ‘*arousal*’ and ‘*valence*’. After further refinement, **PrimEmo** could replace the core decision-making module of a cognitive architecture, such as ACT-R or SOAR. Not only would it confer survival capabilities to the architecture, it would also allow for the possibility of investigating full-fledged emotions, and even emotional expression.

# Table of contents

# List of figures

# List of tables

# Nomenclature

| | |
|---|---|
| 5-HT | Serotonin |
| ACC | Anterior Cingulate Cortex |
| AGI | Artificial General Intelligence |
| AI | Artificial Intelligence |
| BF | Best Frequency |
| BG | Basal Ganglia |
| CA | Central nucleus of the Amygdala |
| CPM | Component Process Model |
| CS | Conditioned Stimulus |
| DA | Dopamine |
| DEAP | Distributed Evolutionary Algorithms in Python |
| DLPFC | Dorso-Lateral Pre-Frontal Cortex |
| EACO | Emotional Appraisal as COherence |

External segment of the Globus Pallidus

| | |
|---|---|
| GPi | Internal segment of the Globus Pallidus |
| GPU | Graphical Processing Unit |
| HRI | Human-Robot Interaction |
| HTTP | HyperText Transfer Protocol |
| JSON | JavaScript Object Notation |
| leaky ReLU | Leaky REctified Linear Unit |
| LHA | Lateral Hypothalamic Area |
| LPFC | Lateral Pre-Frontal Cortex |
| MC | Motor Cortex |
| MGB | Medial Geniculate Body |
| MGm | Medial section of the Medial Geniculate body |
| MGv | The Ventral section of the Medial Geniculate body |
| NAcc | Nucleus Accumbens |
| NEF | Neural Engineering Framework |
| OCC | Ortony, Clore and Collins |
| OFC | Orbito-Frontal Cortex |
| PBWM | The Pre-frontal cortex Basal ganglia Working Memory architecture |
| PFC | Pre-Frontal Cortex |
| PPT | PendunculoPontine Tegmental nucleus |
| PVLV | Primary Value Learned Value |
| REST | REpresentational State Transfer |
| RF | Receptive Field |
| SEC | Stimulus Evaluation Check |

| | |
|---|---|
| SNc | Substantia Nigra pars Compacta |
| SNr | Substantia Nigra pars Reticulata |
| SSC | Somato-Sensory Cortex |
| TD | Temporal Difference |
| UR | Unconditioned or Innate Response |
| URL | Uniform Resource Locator |
| US | Unconditioned Stimulus |
| VMPFC | Ventro-Medial Pre-Frontal Cortex |
| VS | Ventral Striatum |
| VTA | Ventral Tegmental Area |

# CHAPTER 1

---

# Introduction

---

Along with the development of service robots (like Softbank Robotics' pepper robot), smart personal assistants (such as the Google Home or the Amazon Echo) and video games, the need for improving Human-Robot Interaction (HRI) is becoming increasingly important. Within the field of affective science, this new interest for social robots has revitalized two related topics: emotions recognition and emotional expression. Looking at the advances brought by the service and entertainment industries, as well as their products currently on the market, one could easily believe that science already solved the mystery behind human emotions. However, it could not be further from the truth. The reason is that most of the computational models resulting from this endeavor are shallow (Ortony, Clore, & Collins, 1988; Scherer et al., 2010; Sloman, 2001). Shallow here, is taken to mean that the machine executing the program does not have emotions of its own, and neither does it have any understanding of the concept of affect be it for human or virtual agents (a similar argument was made by Searle (1980) concerning the '*mind*' and the machine in his well-known thought experiment called the '*Chinese Room*'). Their shallowness, therefore, comes from the fact that agents in the context of human interaction have no need for emotions of their own. Even without understanding the concept of affect, they are more than capable of categorizing emotions (based on facial expressions for example), then use this labeled information to decide which action to perform next. Therefore, through the clever use of machine learning and other techniques from the field of Artificial Intelligence (AI), it is possible to give the illusion of personality. Consequently, this thesis stems from a desire to go beyond those shallow models and explore the origin and the many roles of emotions within the brain.

## 1.1 Motivation

For most of human history, the concept of emotion was thought to be opposite to that of cognition. While cognition was believed to be the '*cold*' rational part of the mind, emotions were the '*hot*' instinctual or unbridled part. As a result, scientists thought that optimal decision-making should rely solely on cognition. For humans, this meant that emotions had to be restrained or even removed from any mechanism of action selection. In the quest for '*true*' intelligence, many have tried to get rid of any emotion, but to no avail. Recent advances in the fields of neuroscience, cognitive science, as well as affective science, have led to the discovery that emotions and cognition are in fact interdependent systems (Adams et al., 2012; Damasio, 2008; J.-M. Fellous & Arbib, 2005a; Franklin, Madl, D'Mello, & Snaider, 2014; Freitas, Gudwin, & Queiroz, 2005; Goleman, 1995; Ivanovic et al., 2015; Laird, Newell, & Rosenbloom, 1987; LeDoux, 1998; Martinez-Miranda & Aldea, 2005; Oatley & Johnson-laird, 1987; Picard, 2000; Rohrer, 2010; Scherer et al., 2010; Scheutz, 2012; Snaider, McCall, & Franklin, 2011). To put it simply, emotions are required for intelligence to bloom. Thus, trying to get rid of them is futile at best and detrimental at worst. Indeed, for the brain to be able to process the massive flow of sensory information using limited resources, it needs reliable mechanisms to quickly yet accurately assess the situation and trigger an appropriate reaction to maintain the body's homeostasis. Emotions evolved to be one of those mechanisms (Darwin, 1872; J.-M. Fellous & Arbib, 2005a; LeDoux, 2012; MacLean, 1970; Panksepp, 1998, 2005; Rolls, 2007).

Originally, AI and cognitive science had a very similar goal, namely modeling human-level intelligence. While cognitive neuroscientists strive to understand and model the brain mechanisms underlying intelligence, AI researchers use machines to replicate the brain's capabilities. With the rise in computational power, AI systems have become as good as or even better than humans on specific tasks. However, the field nowadays faces yet another challenge. Machines are indeed performing better than humans ever could, but they are limited to a very specific set of tasks. Humans, on the other hand, pursue and maintain many different goals on a daily basis. A small community of researchers with the desire to address this fundamental issue, therefore, split off from the main path adopted by AI. This new branch is nowadays referred to as Artificial General Intelligence (AGI). Using discoveries from related fields, such as cognitive science and machine learning, their goal is to build a computational architecture capable of human-level intelligence (Adams et al., 2012; Goertzel, 2014, 2015). However, for the reasons mentioned above, emotions have notoriously been absent from most projects rooted in either cognitive science or AGI. Resulting in architectures devoid of any emotions. Since it has been established that intelligence requires emotions, it means that neither fields will ever realize its goal. Consequently, before being able to fully understand and model intelligence, it is necessary to investigate the concept of affect and the role of emotions in intelligence.

## 1.2 Problem definition

Recent, so called, cognitive architectures are indeed making an effort to include affect as part of their core design. Even though, researchers have begun to accept the necessity for emotions in order to build intelligent systems, there still lies one more obstacle. Although emotions are ubiquitous and anyone could provide a seemingly endless list of examples, affective science still struggles to define the concept. Throughout affective science's long history, many theories and models have been suggested, but none have been accepted as standard yet. As a result, cognitive architectures usually use the most accepted theory at the time of their creation or a theory reflecting the authors believes (well-known examples can be found in the SOAR (Laird et al., 1987), LIDA (Snaider et al., 2011) and more recently DAC (Verschure & Eds, 2014) architectures). Some even decided against including emotions at all. The most famous in this latter category is ACT-R (J. R. Anderson et al., 2004). It should be noted though, that later research projects have tried to incorporate concepts related to affect within ACT-R, with varying degrees of success. J.-M. Fellous and Arbib (2005b) discuss possible reasons behind this lack of accepted definition, as well as the consequences on the different scientific fields related to emotion. Despite the fact that both cognitive and affective sciences are trying to answer the questions '*what are emotions?*' and '*what role do they play in the context of intelligence?*', the solutions implemented so far do not provide a satisfactory account for the different roles attributed to emotions in intelligence.

For all the reasons expressed so far, this thesis seeks an answer to the question what is '*the role of emotions in autonomous social agents*'? However, before even beginning to tackle this subject, it is necessary to first extract the features differentiating emotions from other sensory and cognitive signals in the brain. This thesis purposefully avoids fully defining '*what emotions are*' (for reasons that will become clearer in Chapter 2), instead relying on its neural underpinnings to understand its role in animals and as a facilitator of intelligence.

## 1.3 Aims and objectives

Although, it was mentioned above that the goal of this thesis was to investigate the role of emotions in autonomous social agents, it is impossible to tackle every aspects of the interaction between emotion and cognition, in a reasonable amount of time. Given that emotions have been said to have evolved to improve an animal's chances in furthering its species (Darwin, 1872; LeDoux, 2012), a more pertinent endeavor would be to analyze '*the role of emotions in increasing the survival potential of social agents*'. Considering, that for living organisms survival is achieved through the detection of relevant opportunities and by acting upon them, it seems only appropriate for this thesis' main topic to focus on the relation between emotion and the decision-making process. Henceforth, the central hypothesis explored is: "*Emotions inform the brain as to the nature of a given situation, interrupt any ongoing action, and guide the decision-making process to increase the chances of survival of virtual agents*". To substantiate this hypothesis, though, the following steps have been identified as necessary:

1. Investigate the current state-of-the-art in terms of theories and models of emotions. The aim is to extract the recurring features for differentiating emotions. As a result, we will be able to discriminate between affective and cognitive systems within the brain, while avoiding the pitfalls associated with having to define the concept of emotion.

2. Among the previously surveyed theories and models, some rely on findings made in neuroscience to explore the origin of emotions in the brain. Adopting a similar approach, the objective is to determine the mechanisms and brain areas supporting emotions. This is facilitated by the characteristics extracted in the previous step.

3. Through the design of a computational model of emotions, the next objective is to show that the mechanisms responsible for emotions also provide increased survival capabilities to virtual agents.

4. Almost as important as emotion, is the topic of decision-making in the animal's brain. As before, the goal is to establish what are the mechanisms at play, as well as the neural underpinnings in charge of deciding the optimal action to perform in any given situation. Fortunately, though it is a more mature field of research that has already provided a standard mathematical model, known as "*reinforcement learning*", along with neural circuits suspected to be involved in this process.

5. A further integration of the architecture from the third step with established models of the reward and punishment systems aims at providing a robust system supporting a

primitive form of emotions. The architecture resulting from this integration is the ideal platform to investigate the role of primitive emotions in the decision-making process. A decision-making mechanism is described as a system managing the access to either a cognitive or a motor resource.

## 1.4 Organization of the thesis

Before exploring the role of emotion in the decision-making process, it is necessary to first understand what exactly is an emotion, or rather in what way is it different from any other mental state? Answering this seemingly simple question, however, is no trivial task. Consequently, this thesis begins by retracing the history of the field of affective science in Chapter 2, and takes a look at the different state-of-the-art models of emotions that have been designed over the last decades. At this point, having developed a better understanding of the different facets of emotions, as well as their neural underpinnings in the animal's brain, we should be able to compile a set of features that will help us distinguish emotion from cognition. Switching to a different topic entirely, Chapter 3 analyzes the neurological mechanisms suspected to be in charge of making decisions in animals. Alongside the biological approach, Chapter 3 also describes the mathematical framework developed to model the decision-making process, which is better known as reinforcement learning.

Next, an anatomically constrained model of fear conditioning, bridging the gap between neuroscience and computer science, is explored in details in Chapter 4. The model of the '*dual pathways*' from the thalamus to the amygdala implemented by Armony et al. (1995) was intended as an early attempt at demonstrating the usefulness of computational models in accelerating discoveries in neuroscience. To validate this model and cement its place in history, the experiment designed by Armony et al. (1995) has been reproduced and the results analyzed in Chapter 4.

Building on the knowledge gathered so far, Chapter 5 introduces **ProtoEmo** a neural architecture inspired by the "*Survival Circuit*" theory (LeDoux, 2012) and hypothesized to be capable of increasing the survival capabilities of virtual agents. In a similar manner, Chapter 6 expands on the information gathered through the **ProtoEmo** architecture, to build **PrimEmo**. This second architecture investigates the role of emotions in the decision-making process. Details of the exact implementation for both **ProtoEmo** and **PrimEmo** architectures are discussed in Appendix A and Appendix B, respectively.

Both architectures designed for this thesis are far from modeling full-fledged emotions. In addition, this thesis only investigates the influence emotions have over a restricted set of cognitive functions, namely decision-making and its associated learning mechanisms. Even though, action selection is a core component of intelligence, it is not the only one. Therefore, the last chapter describes how the current results could be extended and how the **PrimEmo** model could be integrated in a complete cognitive architecture. It also highlights some of the more pressing questions that should be answered before human-level AGI can become a reality.

## 1.5 Contributions

Insofar as I know, the work presented here is one of the first computational modeling projects to explore the dynamic existing between the sub-cortical areas known as: the amygdala, the hypothalamus and the thalamus. While the amygdala is often the center of attention of any study related to emotions, its function is here interpreted in a new context. The resulting system described could be likened to the appraisal mechanism at the core of most cognitive theories of emotions. As is explored in Chapter 2, there are a number of reasons for sub-cortical areas to usually be left to the side when investigating affect. Most research projects that explore the mechanisms underlying emotions focus on higher-level areas, where cognition is more prominent and emotional influence more pronounced as well. Then, there is the fact that for the last few years the field of affective science has been dominated by the cognitive or appraisal approach. As explained in more details in Chapter 2, this is both a blessing and a curse. On the one hand, it has revived people's interest for emotions bringing along new ideas to explore. On the other hand, those new ideas are usually rooted in cognitive science, thus, transforming affect into yet another cognitive system rather than an inter-dependent entity. Finally, due to our current technological limitations the only way to measure human emotions are still through introspection and physiological changes (e.g.: increase in heart rate or skin conductivity). All methods that have been heavily criticized for lacking precision or being biased (LeDoux, 2012; Scherer, 2005). Even functional Magnetic Resonance Imaging techniques are far from ideal to measure and study this phenomenon. Moreover, those methods only measure the outward expression brought forth by emotions, but not emotions themselves.

As a consequence of our inability to clearly define the concept of affect, this project had to work around this problem and instead adopts a perspective that focuses on the functions emotions evolved to fulfill. Even though, the underlying survival circuits introduced in Sub-Section 2.6.2 and modeled in Section 5.1 (but see Appendix A for more details on the exact implementation) ultimately perform cognitive tasks (specifically stimuli classification and decision-making in this case), emotions and emotional behaviors still manage to emerge. For the myriads of theories concentrating on the different aspects of emotions, this project offers an alternative point of view or possible path to a standard definition. As Alexandrov and Sams (2005) best describe it, in this perspective emotion and cognition could be construed as two ends of a continuum. There is no clear delineation between the two concepts. This could be due to the inter-dependence between the two systems, or to the simple fact that emotions really are only a by-product of cognitive activity.

As a matter of fact, a possible interpretation of this thesis' content could be that emotions emerge from the activity of different cognitive systems. This is similar to what was suggested by a movement called '*social constructivism*' (mentioned in Chapter 2). The idea behind social constructivism is that the concept of emotion does not exist in nature and is a product of the mind (Damasio, 2008; LeDoux, 1998; Scherer, 2009). In this thesis, however, far from denying the biological nature of emotions, the idea would be akin to that of Baars' (2005) Global Workspace Theory, in which the spontaneous and synchronous activity of some brain areas give rise to working memory or even consciousness (this view is discussed in more details in Sub-Section 2.7.2). Indeed, LeDoux (2012) indirectly references a similar concept when he places '*global organismic*' states as the raw material from which feelings, and by extension full-fledged emotions, are constructed. As is explained in greater details in Sub-Section 2.6.2, LeDoux (2012) defines a global organismic state as the mental state which results from the detection of a survival relevant stimulus, the activation of the memory, learning, motivational and modulatory systems, the bodily feedbacks, the recruitment of relevant homeostatic and allostatic behaviors, and the subsequent state of generalized arousal. Consequently, the fact that emotions and emotional behaviors did emerge from the computational model developed in this thesis, could be seen as additional evidence toward the idea that emotions labels attributed to recurrent patterns of brain activity.

On a smaller scale, but may be even more important is the fact that **ProtoEmo** is, in my opinion, the missing piece for the Pre-frontal cortex Basal ganglia Working Memory architecture (PBWM), designed by Frank, Loughry, and O'Reilly (2001). The PBWM is a computational model simulating the mechanisms responsible for decision-making in the brain. The architecture can be split into three components, the pre-frontal cortex (PFC), the basal ganglia, and the Primary Value Learned Value (PVLV). Frank et al. (2001) define the PFC as the area where information, or memories, necessary for the completion of the ongoing goal is stored (this is usually referred to as '*working memory*'). The basal ganglia, based on the dopamine signal provided by the PVLV model, decides on which action to perform next (but see Sub-Section 3.3.2 for more details about the gating mechanism implemented by the basal ganglia). In the case of the PBWM architecture, the only available action is whether to update the information stored in the PFC or not, and if so which part of the memory should be updated. The PVLV is a computational model designed by Hazy, Frank, and O'Reilly (2010); O'Reilly et al. (2007) whose purpose is to investigate the processes at play during a classical conditioning experiment. The PVLV model is made of two components: the Primary Value (PV) and the Learned Value (LV) systems. While it is the responsibility of the PV component to learn to activate the dopaminergic system upon receiving a reward, the LV system learns to expect a reward through conditioning (the specificities of the PVLV model are further explored in Sub-Section 3.3.2). According to Hazy et al. (2010); O'Reilly et al. (2007), within the PVLV model the central nucleus of the amygdala is hypothesized to be the excitatory component of the Learned Value (LV) system, whose role is to learn to activate in the presence of a conditioned stimulus reliably associated with an unconditioned stimulus. However, given the scale of the PBWM architecture, the exact mechanisms by which the amygdala is supposed to achieve this function were not fully described. This question is what the **ProtoEmo** architecture strives to answer. Consequently, by extending the PBWM's explanatory power with **ProtoEmo**, the **PrimEmo** architecture (introduced in Chapter 6 and further detailed in Appendix B) forms a complete processing link. From a relevant stimulus entering the brain to the resulting decision, **PrimEmo** is able to explain every step of the process.

To some extent the results and ideas presented in this thesis, also contribute to the field of neuroscience. The amygdala was already known to be an important brain area when it comes to emotions. Whether it be simply processing them or originating them, the amygdala has been the center of attention for quite some time. Most findings, however, come from conditioning experiments. The experimental results of such projects are usually interpreted within the reinforcement learning paradigm. As detailed in Chapter 3, though, reinforcement learning is not directly interested in emotions. There still is the concept of reward or punishment, and more recent research projects have tried to introduce intrinsic features in the computation of the reward signal. This thesis establishes a direct link between reinforcement learning and emotions. Allowing for an alternative interpretation of the empirical data gathered through conditioning experiments. Furthermore, the hope is that including the thalamus and hypothalamus in the design of **ProtoEmo** will spark more interest in these brain areas. Indeed, at the moment the literature concerning those two regions is quite sparse and does not define any underlying mechanisms.

Finally, the results presented in this thesis should also be of interest for the field of Artificial General Intelligence (AGI). Even if, neither **ProtoEmo** nor **PrimEmo** have demonstrated any capabilities in transferring knowledge acquired through one task to solve another (also known as "*transfer learning*"), they are still both examples of control architectures, whose core design relies on emotions. Something that current cognitive and AGI architectures are lacking, as pointed out above. The development of other cognitive component following this model would lead to the establishment of a complete cognitive architecture capable of human-level intelligence.

# Summary

*The fundamental objective shared by both AI and cognitive science is human-level intelligence. Whereas the former is focused on building artificial systems simulating intelligence, the latter seeks to explain the underlying brain systems. For most of human history, emotions were though to be nothing more than disturbances for cognition. Therefore, most studies on intelligence deliberately avoided emotions. As a result, cognitive architectures only partially include emotions in their design, if at all. However, it was recently discovered that emotions and cognition are inter-dependent systems, which means that emotions are required for intelligence to bloom. Furthermore, it was theorized that emotions are a mechanism that evolved to increase an animal's chances of survival. Consequently, this thesis investigates* ‘the role of emotions in increasing the survival potential of social agents*’. Additionally since, for animals, survival is achieved through the detection of relevant opportunities and by acting upon them, this thesis focuses on the interaction between emotions and the decision-making process. When considering action selection as a whole, it is hypothesized that* ‘Emotions inform the brain as to the nature of a given situation, interrupt any ongoing action, and guide the decision-making process to increase the chances of survival of virtual agents*’. While exploring the validity of this hypothesis, the project described hereafter has made some significant contributions. To begin with, insofar as I know this is one of the first computational modeling projects to investigate the interaction between the amygdala, the thalamus, and the hypothalamus. The amygdala was already included in the Pre-frontal cortex Basal ganlia Working Memory (PBWM) architecture as part of the Learned Value component of the Primary Value Learned Value model. However, by implementing representations of the thalamus, hypothalamus, and their connections to the amygdala, in populations of artificial neurons, the **ProtoEmo** architecture completes the PBWM structure. Therefore, the results from the aggregation of the PBWM with **ProtoEmo**, which is referred to as **PrimEmo**, forms a complete processing link, from a relevant stimulus entering the brain to selecting a reaction. Essentially, bridging the gap between emotion and reinforcement learning. Last but not least, the architectures and results presented here should also be of interest to the field of Artificial General Intelligence, since both **ProtoEmo** and to a greater extent **PrimEmo** both implement control mechanisms, whose core design relies on emotions.*

CHAPTER 2

# On emotion: theories and models

Before discussing the feasibility of modeling any affective process, be it emotions, mood or personality, in artificial agents, it is important to understand what emotions are for living breathing animals (humans included) and any other creatures. Therefore, this chapter is dedicated to exploring the concept of affect and the different theories concerning the origin of emotions as seated in the brain.

## 2.1 Passions and the soul

Even though many a scientist considers affective science to be a field still in its infancy, and to their credit it is true that affect as a concept has only recently received a more widespread attention, emotions have puzzled philosophers and all manner of thinkers since at least ancient Greece. Aristotle is usually regarded as one of the first people, building on the work of his mentor Plato, to venture a detailed explanation for the purpose of human emotions and their origin. In his essay entitled '*The Nicomachean Ethics*', Aristotle (350 B.C/1926) explores the concepts of '*pathos*' (a word Aristotle uses to refer to emotion) and describes how the soul when interacting with any object will '*feel*' either pain or pleasure. In turn, those passions influence the way in which the soul and, therefore, the body react to the object. According to Aristotle, it is this very '*feeling*' accompanied by either pain or pleasure, when coming in contact with the external world, that constitutes an emotion. Consequently, for Aristotle emotions are characterized by their valence. Furthermore, similar to the definition adopted by Descartes (1649) centuries later, it is interesting to note that Aristotle assumes a view in which passions are almost indistinguishable from perception. Aristotle then continued his exploration of the '*pathe*' (the plural of pathos) in '*Rhetoric*', where he puts forward two of his most important ideas.

The first, has to do with his categorization of emotions. According to this idea, some of our emotions are '*slow*' enough that we can think about them and influence them before they lead to any action. Those were the deliberate or voluntary emotions. The second category, contains emotions that go by too quickly and are transformed directly into actions, or more precisely in this case, reactions.

The second idea is that emotions and cognition can interact and influence each other. As mentioned, in the case of deliberate emotions one can rationalize and influence the outcome. Aristotle goes as far as saying that it is unethical to let the soul act on a deliberate emotion, without rationalizing about it or influencing it. Indeed, since passions were seen as nothing more than a hindrance for the rational mind, letting deliberate emotions dictate a person's actions would amount to acting without thinking or, more appropriately, acting like an animal.

Although not clearly part of his framework, Aristotle is responsible for setting passion and mind apart from each other. Emotions, in his point of view, are not part of the rational mind they are only a disturbance that should be kept in check, if one is to act morally.

Moving forward in time leads to what is arguably the first ever written theory on emotions. One has to remember that although, Aristotle's treatises touched upon many influential ideas regarding emotions, the primary concerns were on how to behave morally and how to use discourse in the pursuit of enlightenment. On the other hand, Descartes (1649) dedicated his correspondences with Princess Elizabeth of Bohemia to discussing the concept of passions and their relations to the soul.

To better understand Descartes' ideas, it is imperative to first mention the context from which they stem. At the time, the human body was thought to be similar to a machine (centuries later the cognitivists would employ a similar analogy when comparing the brain to a computer, as detailed in Section 2.4). Limbs would move through a system of ropes and pulleys, and sensations were nothing more than pulls and pushes exerted upon the pineal gland. Furthermore, those sensations were transmitted through the veins using the blood and spirits. It is then without surprise that Descartes would theorize about passions using vocabulary borrowed from physics. Given these conditions, Descartes described passion as a vegetative power, a property of something being moved, while action is a moving force, "*like volition in the mind*" (Descartes, 1649). As a matter of fact, the soul was separated from the body and made of the individual's thoughts. Those thoughts, according to Descartes (1649) were of two kinds: '*action*' and '*passion*'. On the one hand, the rational soul via the pineal gland '*acts*' upon the body to move it according to our will. Reciprocally, the body sends spirits through the pineal gland to influence the soul so that the individual can sense the world. It should be noted that although the soul is attached to the body in many locations, Descartes (1649) describes the pineal gland as the main interface through which the soul and the body most efficiently influence each other. Consequently, actions come from and fully depend on the soul. Passions, though, are more complex. In the most general sense passion can be interpreted as perception, since it is made of anything new that arrives to the soul from the body (or more specifically from the external world via the sensory devices of the body). A stricter definition sees the passion described as feelings or emotions which result from the agitation of the spirits and that move the soul. Furthermore, Descartes (1649) defines the functions of the passions as follow:

> "*. . . l'usage de toutes les passions consiste en cela seul qu'elles disposent l'âme à vouloir les choses que la nature dicte nous être utiles et à persister en cette volonté;comme aussi la même agitation des esprits, qui a coutume de les causer, dispose le corps aux mouvements qui servent à l'exécution de ces choses.*"
> — Descartes (1649, p. 343)
>
> [*The principal effect of all the human passions is that they move and dispose the soul to want the things for which they prepare the body. Thus the feeling of fear move the soul to want to flee, that of courage to want to fight.*]

In modern terms this definition would equate the passions with motivators as described by Rolls (2013). In his discussion, Descartes clearly specifies that the '*things*' for which the soul and body are being prepared is "*that which nature deems useful for us*". Thus, suggesting, that emotions (or in this case passions) are an innate mechanism that evolved to help humans evaluate the objects with which they interact in terms of their usefulness for survival. Two centuries later, this idea would be substantiated and expanded to the whole animal kingdom, by Darwin (1872). Furthermore, on the one hand, since passions can only influence the soul and the soul acts mostly through the pineal gland, this marks the first time that emotions were located in the brain, as opposed to the guts, heart or an ethereal entity (Damasio, 1997, 2008). On the other hand, on the contrary to what Aristotle thought, Descartes (1649) considers that truthfully action and passions are two names given to the same concept. The only reason he sees for those labels to exist is a difference in point of view. The same event is referred to as action from the perspective of the person originating said event, while it is described as passion in the context of the person on the receiving end. Finally, it is also interesting to note that Descartes, defines the existence of six '*primary passions*'. All other emotions result from the combination of those primary six or are sub-species. Another idea reminiscent of Aristotle's '*slow*' and '*fast*' emotions.

## 2.2 The birth of affective science

The year 1872, marks the most important turning point in the history of affective science. After 34 years of research on emotions, Charles Darwin (1872) published his now famous book entitled '*The Expression of the Emotions in Man and Animals*', in which he introduced some of the most revolutionizing ideas. Analyzing the many portraits he accumulated, Darwin (1872) noticed that there were some peculiar similarities in the ways people and animals expressed certain emotions. Building on his principle of '*survival of the fittest*', which he introduced to the world a few years prior (Darwin, 1859), Darwin theorized that at some point in the past one of our animal ancestors evolved a primitive form of emotions. Using '*fear*', for example, this common ancestor would have been able to predict the dangerous nature of a given situation, rather than react to the appearance of a predator. Given that this new '*trait*' increased its chances of survival, according to the theory of evolution it should have been passed down from this common ancestor to the next generation, mutating until it became what humans refer to as emotions. The important consequence of this theory is that emotions are not a property unique to humans anymore. Indeed, it implies that humans share some of their emotions with other animals, and that there is a set of '*basic*' emotions common to all humans. Thus, upholding the tradition of categorizing emotions as did both Aristotle and Descartes before. Furthermore, if emotions really evolved from a common ancestor, it follows that the fundamental brain structures originating emotions should be preserved between species as well (Ekman, 1973). Therefore, Darwin's (1872) discovery entails that any empirical results from animal experiments could be used to draw conclusions about human emotions. In addition to defining emotions as a mechanism to improve survival, Darwin (1872) characterizes their effect on the brain and the individual in general as being either exciting or depressing. Darwin (1872) cite as examples of exciting emotions, anger and joy. Indeed, in both emotions lead to energetic movements and a higher heart rate. On the contrary, sorrow and grief are depressing emotions, since they bring about almost lethargic movements, lack of motivation, and slower cognitive processing. As it will become clear in the remainder of this chapter and the next one, Darwin's ideas will be a source of inspiration for many theories and models of emotions, even to this day.

Ten years after Darwin's discoveries, William James (1884) published his seminal paper in which he set out to answer once and for all the question '*What is an Emotion?*' and how they are produced in the brain. Following James' theory, and similar ideas that would be put to writings by Lange (1885) a year later, emotions are a result of the bodily changes, occurring in reaction to an event, being fed back to and interpreted by the brain. A more palatable way to put it is to say that, people do not weep because they are sad and scream because they are fearful. In fact, according to the James-Lange theory, it is the contrary: we are sad because we weep and are fearful because we scream. If someone was to meet a bear while walking through the forest, this encounter would trigger a host of reactions inside his body, such as increasing heart rate, release of adrenaline and increase of skin conductance, as well as outside, like running away for example. The experience of these internal and external bodily changes is what constitute the emotion of fear. It stems from this definition that if one is to have emotions, then one needs to have a fully functioning body to trigger them. Furthermore, even though James (1884) recognizes that their are feelings of pleasure and pain, or excitement and depression, associated with certain '*nerve-actions*', because those mental states are not associated with any distinct bodily expressions, he does not consider them as '*standard*' emotions.

It is also interesting to note that to the question of how does the body knows how to react in each situation, James explains that part of the human brain (if not its entirety) is made of a hard-wired structure which maps sets of sensory stimuli to bodily reactions, or as he elegantly describes it:

> "*To begin with, readers of this Journal do not need to be reminded that the nervous system of every living thing is but a bundle of predispositions to react in particular ways upon the contact of particular features of the environment.*"
> — James (1884, p. 190)

As history and the rest of this chapter shows, James' controversial ideas were the sparks that would later give birth to the field of affective science as it is known today. Even though, his ideas would remain unchallenged for a long time.

## 2.3 A modern history

Through a series of experiments involving animals whose cortical connections had been severed, Cannon and Bard (Cannon, 1927, 1929) were able to show that emotions could exist without a body proper, thus ending the fifty years reign of the James-Lange theory of emotions in psychology and, at the same time, being the first to make practical use of Darwin's discovery: that of animal emotions as homologues to human ones.

More specifically, Cannon challenged the James-Lange theory on five points:

- First in 1904, Woodworth and Sherrington (1904) experimented on dogs, whose spinal cord and vagus nerve were transected, so as to remove any connection between the brain and the organs that were believed to originate the sensory feedback required for the creation of emotions. Then twenty years later, Cannon, Lewis, and Britton (1927), were also able to keep alive cats whose sympathetic division of the autonomic nervous system was removed, effectively making them unable to both react to particular events, as well as feel any of the bodily changes that could have occurred. In both experiments, no changes in emotional behaviors were observed, therefore refuting the concept of embodied emotions present in the James-Lange theory.

- In another set of experiments conducted by Marañon (Cannon, 1927), patients were injected with a dose of adrenalin, which in the body produces changes similar to that of intense emotions, such as constriction of the blood vessels, liberation of sugar by the liver and interruption of gastrointestinal functions. The experiment thus designed was supposed to show that by artificially re-creating the symptoms of an intense emotion, the injected patients would react accordingly and express their emotions. Instead, when interviewed about their current emotional state, by Marañon, most of the patients would say that they felt, and indeed expressed, no particular change in emotions. Interestingly, in some cases subjects would describe their emotional state as: "*I feel as if afraid*" or "*as if awaiting a great joy*". The results of Marañon's study again opposed what the James-Lange theory predicted.

- Central to the James-Lange theory are the viscera and vasomotor system. According to James' ideas, feedback from changes in the viscera are the main information used to discriminate between different emotions. However, in Cannon's point of view, relying on the knowledge available at the time, the sympathetic division of the autonomic nervous system projects too diffusely to the different glands and smooth muscles throughout the body, to produce a specific pattern of activation, that one could use to discriminate between emotions. As a result bodily changes that are observable for a feverish person (a non-emotional state), are also present for a person experiencing happiness and fear (Cannon, 1927, 1929).

- Furthermore, Cannon argues that smooth muscles and glands, on the contrary to striate muscles used to move limbs, respond to signals coming down from the brain with a certain sluggishness. For example, in animals like cats, dogs, and horses, the latent period of smooth muscles, which is the amount of time between the brain sending a signal and the muscle beginning to move in reaction to that signal, is in the range 0.25 seconds up to 0.85 seconds. And for salivary glands in dogs, Pavlov (Cannon, 1927) measured a latent period of six minutes. In the realm of humans, Wells, Forbes, and Whitley (1912) measured a latent period of about three seconds for the psychogalvanic reflex. In another study conducted by Wells (1925), affective reactions to pictures of men and women had a latent period of 0.8 seconds on average. If the James-Lange theory is to be believed and emotions are really born "*from reverberations from the viscera*" (James, 1884), it would take seconds for an emotion to emerge out of a particular situation, the world would go by too quickly for us to react to it.

- Finally, contrary to popular beliefs dictating that the deeper you travel inside the body, the more sensitive organs are, surgeons discovered, around the time of Cannon's writing, that the alimentary tract of unanesthetized patients could be cut, torn, crushed or even burnt without any feeling of discomfort. Arguing that the same can apply to both the rest of the viscera and blood vessels, Cannon concludes that those organs are not sensitive enough to provide the basis for the creation of emotions (Cannon, 1927, 1929). If only damage of an extreme nature to the viscera triggers a sensation, how can they originate subtle changes in our emotional state?

Even though the weaknesses highlighted by Cannon and Bard were well founded and had ample support from other studies, they would in turn receive their fare share of criticisms, leading to Cannon suggesting his own theory on emotions and their associated brain mechanisms.

In an experiment conducted in 1925, Cannon et al. (1927) observed that a slight stimulation would trigger fits of intense rage in cats whose cortical connections had been severed. Cannon would later call this phenomenon "*sham rage*" since the animals involved in the study had neither a physical target toward which they could direct their anger, nor any concrete reason to be angry in the first place. He deduced that if the brain had an emotion center then it must be confined to the sub-cortical structures. Furthermore, removal of the thalamus in the same cats would render the animal incapable of expressing any emotion. Hence, Cannon concluded that the main center for the brain's emotional system was the thalamus. Moreover, he described the interaction between the cortex and the thalamus as being inhibiting. The reason why the cats were displaying intense emotions while their thalamus remained, and the cortex was removed, is because the thalamus is the center of a whirlpool of emotions. The role of the cortex according to Cannon is to modulate the emotional activity that is constantly happening within the thalamus. Therefore, by removing the cat's cortex, Cannon was effectively removing any emotional restraint the animal had. Lesioning the thalamus prevented the animal from having any emotion in the first place. From those observations, Cannon et al. (1927) theorized that "*the peculiar quality of the emotion is added to simple sensation when the thalamic processes are roused*" (Cannon et al., 1927, p. 120). Both theory and observations clearly point in the direction of emotions having an exciting or depressing effect on the brain, and, therefore, any of its ongoing cognitive activities.

In a later paper, Cannon (1931) contributes more details to the mechanisms underlying his theory of emotions. On the contrary to what one might think, given Cannon's previous complaints, he acknowledges the fact that James' ideas were heading in the right direction, however due to either a lack of knowledge or technological limitations, the James-Lange theory completely overlooked the fundamental role of the thalamus in the communication between brain and body. Based on the evidence that any signal going in or out of the cortex had to pass through the thalamic region, Cannon described it as the "*coordination center for the so-called 'emotional' reactions*", hence grounding its position as the main component in his theory.

The similarities between the James-Lange and Cannon-Bard theories become clear when illustrated by the diagrams of figure 2.1. It should also be noted that the Cannon-Bard theory does not separate the body from emotions, since sensory data is still required to first trigger both the bodily reactions and the exchange between thalamus and cortex, that will lead to the creation of the emotion.

In 1962, influenced by the James-Lange theory, but very well aware of the criticisms made by Cannon et al. (1927), Schachter and Singer (1962) suggested the '*Two-Factors*' theory according to which emotion is a function of the state of physiological arousal accompanied by an appropriate cognition. Thus, the physiological arousal is the driving force, while cognition has more of a steering function. To give a similar example as the one employed for

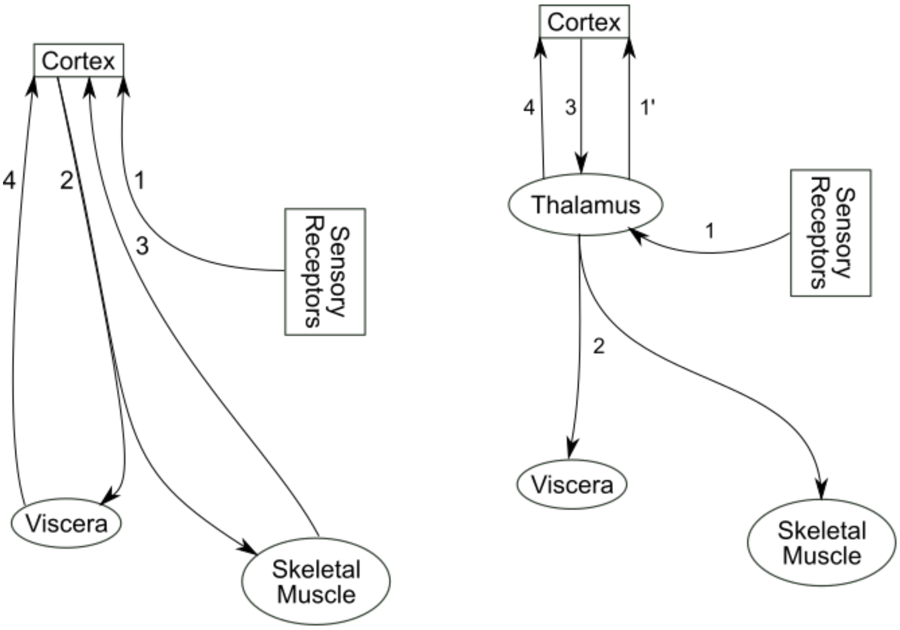


Figure 2.1 *This figure depicts the circuits responsible for emotions and emotional expression, as theorized by James (1884) (on the left) and Cannon (1931) (on the right). Already the differences, in opinion are clearly visible, as the Cannon-Bard model places the thalamus at the center of the circuit. The James-Lange theory, on the other hand, heavily rely on the cortex to process sensory and feedback information, as well as emotionally react. Following the numbers, in order, depicts the sequence triggered by incoming sensory stimuli and leading to emotion and its expression.*

the James-Lange theory of a person walking around in the woods suddenly coming across a bear. The perceived situation of '*proximity of a bear*' would initiate a state of physiological arousal. This new physiological state would then be interpreted or contextualized using the person's knowledge of bears and how ferocious one can be when startled. Finally, leading to the physiological state of arousal being labeled as '*fear*'.

To substantiate their hypothesis, Schachter and Singer designed a clever experiment, in which participants would be either injected with epinephrine or a simple saline solution. Epinephrine would induce in the subject an artificial state of arousal, while the saline solution was considered to be the control case in which physiological arousal was absent, or negligible. During the injection procedure, half of the participants receiving a dose of epinephrine, were given a detailed explanation of the effects of the substance on the human body. The other half did not get any explanation, leaving them in a state of arousal with no appropriate cognition. After injection, irrelevant of the substance, the participants were left in a room with an actor. The actor could play one of two roles, either that of another happy participant, or that of another angry subject. The actor was introduced to steer the emotions of the participants that were injected with epinephrine, but given no explanation about its effects. By providing an adequate framework, the subject would finally be able to label his state of arousal. Following Schachter and Singer's hypothesis it was expected that subjects injected with a saline solution (no physiological arousal) or injected with epinephrine and knew of its effects (physiological arousal and appropriate cognition), should not feel any particular changes in their emotional state. While subjects injected with epinephrine (physiological arousal), but that did not have any clues as to its effects (no appropriate cognition), would label their new arousal state as either '*happiness*' or '*anger*' depending on which actor they had been paired with.

The experimental results proving Schachter and Singer right, the two-factor theory was published, hence marking the first point in history where cognition and emotions, or passions and reason, were brought together. However, the theory's impact on the history of affective science is much greater. Introducing cognition as a main component in the emotion machine meant that cognitive scientists, who until then had kept their distances, could join and apply their knowledge to the problem, breathing new life into the research on emotions.

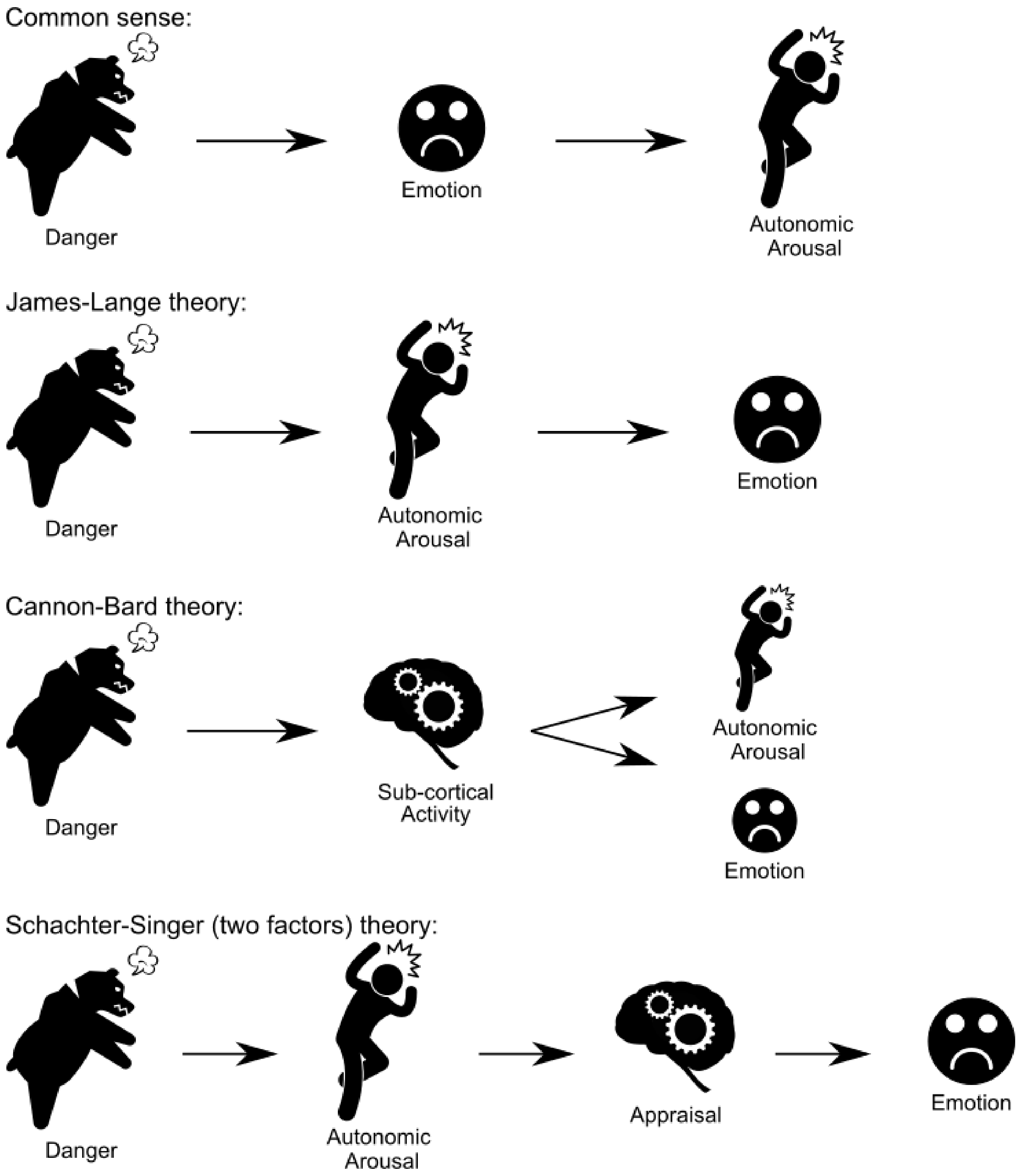


Figure 2.2 *Starting from the same triggering event, of a person suddenly encountering a bear, this picture highlights the main steps of an emotional episode, as depicted by different theories. From top to bottom, common sense would have us think that running away (autonomic arousal) is a consequence of being afraid (emotion). The James-Lange theory, on the contrary argues that emotion is a consequence of our reaction. Cannon and Bard in the meantime, place an emphasis on sub-cortical activity (especially the thalamus), as the main system responsible for both autonomic arousal and emotion. Finally, Schachter and Singer introduce the concept of appraisal. Given a situation and a state of arousal, the appraisal process produces an emotion (a simple label for the state of arousal).*

## 2.4 Appraisal theories

Throughout the first half of the twentieth century, the field of psychology was dominated by the behaviorists (LeDoux, 1998). Breaking the tradition instigated by Descartes' (1637) '*Cogito, ergo sum*', the researchers part of this new movement thought that psychology should be the study of objective and observable facts. Since consciousness and emotions are notoriously subjective and unmeasurable, they could not possibly be studied scientifically. Thus, consciousness and the mind in general, which came to be known as '*ghost in the machine*' ( LeDoux (1998), but see also Ryle (1949) for the original philosophical definition of this concept), were ostracized from the field of psychology. The behaviorists went as far as ridiculing anyone who dared to mention those mental states.

Ironically, it is from within this oppressive environment that one of the most important and for the time the broadest theory of emotions was written. After spending almost ten years studying emotions in relation to personality and mood, Arnold wrote, what is nowadays considered her magnum opus, a two volumes book entitled '*Emotion and Personality*' (Arnold, 1961). In it, realizing that the prominent theories concerning the emotional system all had a fundamental flaw, she fleshed out the concept of '*cognitive appraisal*' (Reisenzein, 2006).

The theories described so far focus on the experience and expression of emotions, while assuming the following general definition for an emotional episode:

- First, a stimulus reaches the brain.
- This stimulus triggers a reaction or some sort of bodily changes.
- The sensations related to those bodily changes are fed back to the brain.
- Finally, some mechanism based on the information available give rise to an emotional experience and/or expression (depending on the theory adopted this process can involve more than one step).

However, none of the theories give an exact account of the brain's mechanisms to differentiate between a stimulus that is cause for a reaction and other stimuli that do not engender any bodily changes. As an answer to this problem, Arnold introduced the concept of cognitive appraisal as:

> "*What is perceived does not lead to emotion unless it is first appraised as good or bad for us here and now, requiring this or that action. The object itself may*

> *be present or absent, remembered or imagined: as soon as we appraise it, it has affected us personally in some way, it leads to emotion. This appraisal is not a rational judgment, it is immediate, intuitive and as inevitable as sense perception itself. . . . This appraisal can be called a sense judgment, to distinguish it from reflective judgment. It immediately follows and completes perception, hence can be known as a separate process only when we come to reflect upon it.*
> *Usually, the sequence perception-appraisal-action is so closely knit that our everyday experience's never the strictly objective knowledge of a thing, it is always a knowing-and-liking, or a knowing-and-not-caring. There is hardly anything we simply note without appraising it at the same time. Whether a thing is pleasing to the eye, delectable to the palate, useful as a tool, a treasure to possess, it is never simply there, apart from us, it is always seen in some relationship to ourselves. In fact, strictly objective observation has to be carefully taught and meticulously acquired to make so-called scientific method possible at all.*" — Arnold (1959, p. 407)

From this broad definition of appraisal some important properties concerning the sequence from perception to emotion, as well as the different concepts involved, can be extracted. First and foremost is the fact that appraisal, similar to the Schachter-Singer theory, does not directly bring about any behavioral reaction unless the emotion experienced is of great intensity. Instead, the appraisal of a given situation leads to the potentiation of appropriate action tendencies to prepare the body for dealing with said situation. The output of the appraisal process as defined by Arnold is simply an '*action tendency*', which will engender an emotional experience and influence the decision-making process, once it reaches consciousness. Second, the sequence of events that form the beginning of the emotional episode (the first step in the previous global definition), is clearly established as perception, then appraisal and finally action. Meaning that for an individual to experience emotion, he will have to first make sense of his surroundings, extracting the different objects. Only then can the appraisal process match the situation to the person's beliefs, desires and coping potential, and influence his actions. Similarly to the Cannon-Bard theory this sequence implies that emotional experience is not bound to any bodily changes. It only relies on the output of the appraisal mechanism. Finally, appraisal is defined as an automatic and unconscious process indistinguishable from sense perception until it reaches consciousness. Where from a person will have access to the object target of the appraisal system, hence, lending support for introspection as a legitimate method for scientifically studying human emotions in psychology.

Although, Arnold has received a lot of attention from the cognitive science community, for her contribution and ideas, most of her work has remained in the shadow, unappreciated. Her books, not only introduced the concept of appraisal, they reached further and described the brain circuits supporting the appraisal mechanism and the whole emotion system. Furthermore, she also drew conclusions as to the repercussion her theory would have on psychology and the study of personality. Arnold justifies this quest for a general theory of emotion by saying that:

> *"If we analyze this process from perception to emotion and action, we shall find that we have to complete our analysis by recourse to neurological research. Emotional expression, autonomic changes and overt action imply functioning circuits in the brain; hence no explanation can be complete unless we can show what is happening in the brain during emotion."* — Arnold (1959, p.407)

Even thought Arnold is recognized as the author of the appraisal theory, the scientist usually credited for making it popular as well as scientifically acceptable for researchers in the 1960s (keep in mind that the behaviorists, even if weakening, still had a strong grip over psychology) is Richard Lazarus (LeDoux, 1998). Borrowing Arnold's ideas (Reisenzein, 2006), Lazarus (1966) set out to prove that a person's reaction to (or coping with) a stressful situation can be influenced by manipulating his perception of said situation. According to the appraisal theory, depending on how the situation is perceived the person should experience different emotions.

With this very goal in mind, Lazarus designed an experiment in which subjects had to watch a short, but rather gory film of a circumcision ritual involving teenage males from an aboriginal tribe in Australia. For a third of the participants, the soundtrack accompanying the film was left intact and contained all the gruesome details. For the other two thirds, the soundtrack was either minimized (lower volume) or intellectualized. The experimental results showed that the participants who were watching the film and had the soundtrack with the gruesome details at full intensity, had higher activations of the autonomic nervous system (ANS) and felt worse after the experiment, when compared to the other two groups. In line with the appraisal theory, Lazarus concluded that the soundtrack accounts for a difference in the appraisal pattern, thus, giving rise to distinct emotional experiences among the three groups of participants.

Lazarus' theory on stress emotions and his extensive experimental work providing empirical evidence for the predictions made by Arnold's appraisal theory is the main reason for him to be considered a pioneer in this subfield of affective science and in psychology in general.

By 1980, appraisal theory had grown in popularity within the field of psychology attracting the interest of many cognitive scientists. However, this began to change after Zajonc (1980) published his paper entitled '*Feeling and thinking: Preferences need no inferences*'. There he described a series of cleverly designed experiments based on the mere exposure effect (Zajonc, 1968). During one such experiment, participants were exposed to a particular Japanese ideogram for a short period of time. They were then asked to choose among a set of ideograms which one they preferred. The results showed that the subjects reliably chose the ideogram that they had been pre-exposed to. The clever part was to present the stimuli subliminally, long enough to form a preference but short enough for the subject to be incapable of recognizing the ideogram. From the results of this and many other experiments that Zajonc and his colleagues performed, he concluded that emotional reaction was not preceded by perception. Zajonc (1980) complemented this conclusion by arguing that:

> "*Before we evolved language and our cognitive capacities, which are deeply dependent on language, it was the affective system alone upon which the organism relied for its adaptation. ...Thus, if the most recent version of homo sapiens specifies that affective reactions are mediated by prior cognitive processes — as contemporary cognitive views would have it — then at some point in the course of evolution, affect must have lost its autonomy and acquired an intermediary in the form of cold cognition. This scenario seems most unlikely. When nature has a direct and autonomous mechanism that functions efficiently ... it does not make it indirect and entirely dependent on a newly evolved function.*" — Zajonc (1980, p. 170)

Even though, Zajonc (1980) describes cognition as "*deeply dependent on language*", which is not the case and could be argued limits the impact of his experiment, this argument still supports the idea that cognition does not necessarily precede emotions. In fact Zajonc (1980) goes even further and states that emotions and cognition are two separate systems. It is interesting to note that Zajonc (1980) uses a reasoning similar to Darwin's (1872) to illustrate his point. Invoking the principles on which evolution is theorized to operate, to emphasize the incongruity of the (at the time) standard information-processing model of emotion, which emphasizes the reliance of emotion on cognition and, therefore, diminishes the distinction between the two concepts. Zajonc's discoveries, however, did not spell the downfall of the appraisal theory, but it served as a reminder for later researchers that emotion is not a simple cognition.

The middle of the twentieth century saw the development of the computer and with it the emergence of artificial intelligence (AI), as a field seeking to reproduce the capabilities of the human mind through computer simulations. Scientists quickly realized the similarities in the way a computer processes information and the inner workings of the brain, a point of view reminiscent of Descartes (1649) own approach to the human body as a machine. Inspired by this new perspective and the recent findings in cognitive science, psychology saw the birth of a new movement, that would later replace the behaviorists, called '*functionalism*'. In the point of view of functionalism, a computer processing the expression $2+5$ is using the same underlying strategy than a human brain, to determine that the answer is 7. As a result of this decoupling between a process' implementation and its underlying hardware, it was no longer necessary to understand the intricacies of the brain to describe its functioning. If it was possible to build a computer simulation mimicking the output of a human brain, then the algorithm used by the simulation was considered a viable explanation for how the brain processed information to reach the same conclusion. In short, according to the functionalism doctrine: "*the mind is to the brain as a computer program is to the computer hardware*" (LeDoux, 1998). Contrasting with the behaviorists, functionalism finally saw the return of the mind in the brain. However, the mind was still not present in its complete form, since the functionalists focused on only the unconscious processes of the brain. Therefore, consciousness was once again left out and with it the possibility of expressing emotions. Thanks to Arnold's appraisal theory, though, emotion was not completely absent from the brain anymore, since appraisal had been described as "*immediate, intuitive and as inevitable as sense perception itself*" (Arnold, 1959, p. 407).

Even after Zajonc's discoveries and partly due to AI and the functionalists, appraisal theory continued to flourish from the 1960s to this day. As a matter of fact, by the 1980s within the field of psychology appraisal theories were about the only approach to emotion (LeDoux, 1998, p. 53). LeDoux (1998) even considers this approach as the one that almost got it right. The only complaint he puts forth is related to the over-reliance of appraisal theories on cognitive processes to explain emotion, hence, making it hard to distinguish emotion from cognition. Additionally, Marsella, Gratch, and Petta (2010, p. 27) describe appraisal theories as "*currently a predominant force among psychological perspectives on emotion and arguably the most fruitful source for those interested in the design of symbolic AI systems, ...*", therefore, echoing LeDoux's (1998) claim, as well as extending it to the field of AI and affective computing. As a result of its close relationship with computer science, as well as psychology and cognitive science, it seems like appraisal is the ideal theory for implementing cognitive architectures endowed with emotions and other models of emotions used in the entertainment and service industry. At the same time its attractiveness for experts in the field of cognitive science has not diminished, which also explains its influence as a sub-field of affective science.

Before analyzing the concept of affect through the lens of statistics, as well as exploring the possible neural substrates originating emotions and feelings in the brain of animals, however, the next Sub-Section presents the component process model which expands the appraisal theory introduced by Arnold (1959, 1961). The advent of computer science, and later robotics, has allowed researchers to apply their knowledge to real world problems, as well as test the theories produced by the different approaches to emotions in a safe and controlled environment. In contrast to the theories described so far, which provide abstract mechanisms to explain emotions and might neglect some aspects or leave some questions open for others to answer, computational models are a realization of those theories. As such, computational models need to specify every detail before an implementation is possible, and then justify the values assigned to its many parameters once embodied or implemented. Since they force a single interpretation of their supporting theories, computational models of emotions stand as an expression of the author's point of view. In addition, the many design decisions made during the building process emphasize the theory's fundamental components, and provide an indication of the direction towards which the whole field is heading.

### 2.4.1 The Component Process Model

Introduced by Scherer (1984), the Component Process Model (CPM) is a framework detailing the appraisal process and its interaction with other cognitive subsystems. The CPM has been under development ever since its inception (Sander et al., 2005; Scherer, 2009; Scherer et al., 2010; Scherer, Schorr, & Johnstone, 2001), but has never been fully implemented. The reasons for this lack of implementation will become clear later in the description.

As a general concept, Scherer considers emotions to be a mechanism that evolved to replace instincts, usually conceptualized as rigid stimulus-response chains. In this perspective, emotions are a flexible mechanism, which by decoupling stimulus and response, allows adaptation to dynamic and complex environments. Hence, enhancing the survival capabilities of an agent endowed with such a structure. In the context of the CPM, however, Scherer defines emotions as:

> *"An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems. It is a response to the evaluation of an external*

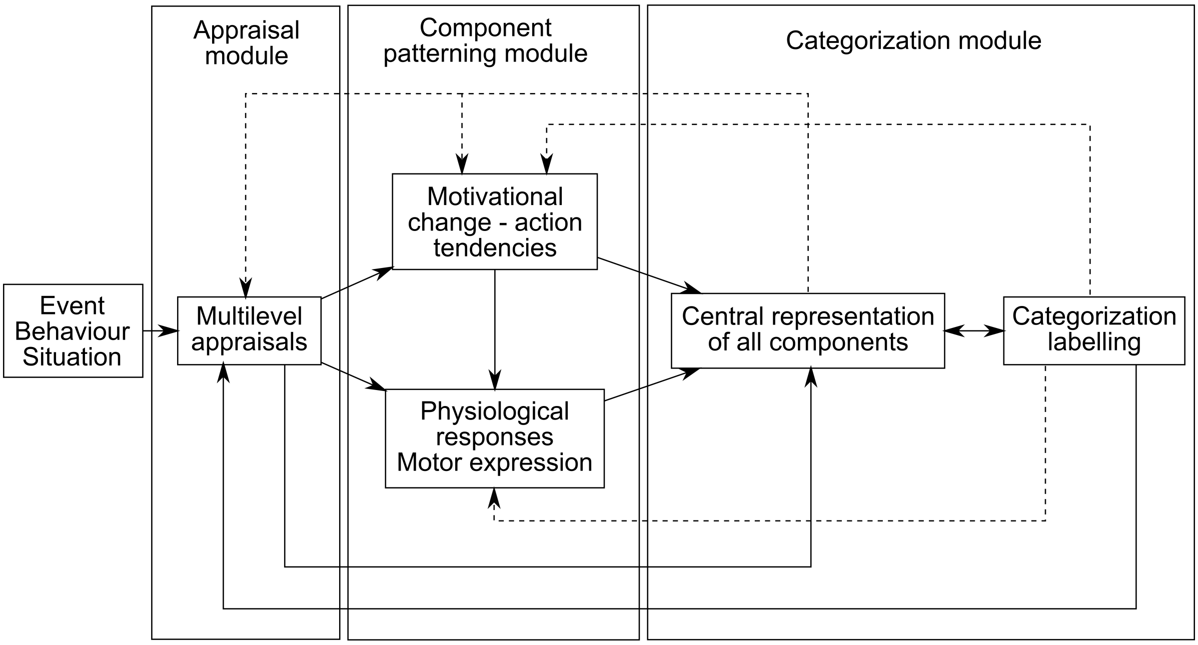


Figure 2.3 *General architecture of the component process model of emotion.*
Source: Scherer et al. (2010, p.50), used with permission from Oxford Publishing Limited.

> *or internal stimulus event relevant to major concerns of the organism.*" — Scherer et al. (2010, p. )

The five '*organismic subsystems*' mentioned in the previous definition, each fulfill a specific function among: *1*) evaluation of objects and events; *2*) regulation of internal subsystems; *3*) preparation for action; *4*) signaling of behavioral intention; *5*) monitoring of internal state and external environment. It is important to note that Scherer adopts the idea of emotion as a multi-componential construct, which is reminiscent of previous theories (Cannon, 1927; MacLean, 1952; Papez, 1937). Indeed, although not the focus of his theory, MacLean (1949) also suggested the existence of '*cerebral analyzers*' receiving a wide variety of internal and external stimuli. Emotions then were the results of patterns of activity in those analyzers. However, this is not a view that every scientist shares even today (Dalgleish, 2004; Damasio, 2008).

The overall architecture of the CPM framework, as suggested by Scherer, can be split into three principal modules (shown in Figure 2.3): *1*) the appraisal module; *2*) the component patterning module; *3*) and the categorization module.

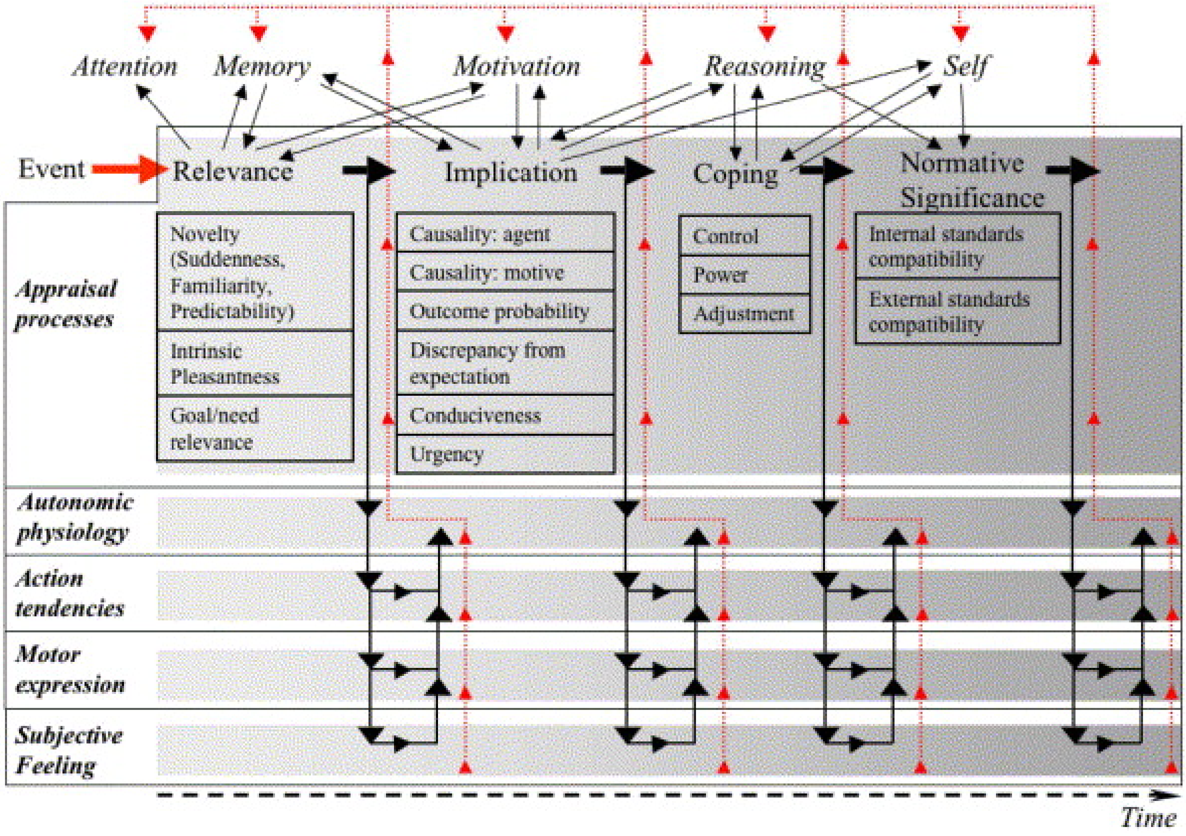


Figure 2.4 *A comprehensive illustration of the component process model of emotion.* Source: Sander et al. (2005, p.321), reproduced with permission from Elsevier.

### Appraisal module

The appraisal module, is the most important part of the CPM framework, as its outcome indicates if a stimulus is to trigger an emotional episode. As such, the results of its processing components, drive the agent's reaction to a particular stimulus event, and are major components of the integrated representation that characterizes the subjective emotional experience, better known as '*conscious feeling*'. According to Scherer, to ensure that an agent's reactions are adaptive and favor its survival, four global appraisal objectives need to be reached. Each of them can be further broken down into sets of '*stimulus evaluation checks*' (SEC), each evaluating a distinct property of the stimulus event.

In order to successfully survive in a dynamic and complex environment, every living creature needs to remain aware of the different threats surrounding it, as well as any potential opportunity that might improve its situation. However, to perform this constant monitoring task animals have only access to very limited resources, in terms of energy supplies and also mental capabilities. Therefore, Scherer argues that the first of the appraisal objectives, is a low-level process sorting incoming stimulus events according to their relevance for the agent's needs and goals. On the one hand, depending on its relevance, a stimulus could be completely ignored, thus saving precious resources from being wasted. On the other hand, relevant stimuli will be attended to in order of their importance. Extreme cases, such as threats to an agent's survival, will interrupt any ongoing processing and enlist the agent's full attention and mental capabilities (the main reason for scientists to consider emotion to be an alarm system). To perform this low-level filtering, a stimulus event is evaluated against three criteria: *1*) '*novelty*', which is a measure of how abruptly a change in the environment occurred (a fast moving object or a sudden burst of noise) and of how familiar or predictable a given event is, based on previous experience; *2*) '*intrinsic pleasantness*', judges the likelihood of an event resulting in a state of pain (to be avoided) or pleasure; *3*) finally, the agent's goals and needs are taken into account to grade the pertinence of the stimulus event (abbreviated in the CPM as '*relevance to goals and needs*'). For Scherer, the agent's goals and needs are fundamental components both in the context of the CPM and for animals in general. Arguing that "*without needs or goals, no real emotions*". As a consequence, if virtual agents are to be given emotions, the concepts of needs and goals have to be implemented first. Two requirements that Scherer considers to be fairly easy to fulfill (Scherer et al., 2010).

Once a stimulus event is deemed relevant for the agent, a plan has to be formulated to either approach or avoid it, depending on its intrinsic pleasantness. In Scherer's opinion, five checks are minimally required for the animal to learn more about the stimulus itself, its cause and probable consequences. The checks suggested, to extract the required information, are: *1*) ‘*cause*’, one of the most important check (and also difficult to implement for virtual agents), since it attributes agency and intention; *2*) ‘*probable outcome*’, for which an individual has to consider every possible outcome and compute its likelihood of happening given the sequence of events leading to the current stimulus, as well as past experience; *3*) in an effort to preserve resources and to increase the speed with which an animal takes decision, its brain usually learns repeatedly occurring sequences of events, thus a stimulus “*failing to meet such expectations*”, might be cause for concerns; *4*) ‘*conduciveness*’, is an assessment of the degree to which a given event hinders or facilitates the realization of active goals and needs; *5*) and ‘*urgency*” provides an estimate for the importance of an event and for the speed at which a reaction has to be produced.

However, animals are capable of more than monitoring changes in the environment. Rather than being swayed around, they can act upon the world. Modifying the environment until it meets their needs. In the context of the CPM framework, this possibility is expressed through the third appraisal objective: coping potential. If an animal sees an opportunity to gather more food, it stands to reason to pay attention to stimuli predicting it and even expend some energy to reach said food resource. On the contrary, if an event, like a natural disaster, cannot be influenced, an animal might try to adapt its current strategy to accommodate the consequences. In the worst case, it will simply be overcome with a feeling of helplessness. To estimate the potential an agent has to cope with a given stimulus, the CPM relies on three checks: *1*) ‘*control*’ and *2*) ‘*power*’, are usually used interchangeably by authors, however Scherer makes the distinction here between control, as a property describing how susceptible to external influence an event is, and power, as a property of the agent defining its capabilities, in terms of money, strength or knowledge, to control a situation; *3*) in case an agent has no mean of impacting an event’s outcome, its ‘*potential for adjustment*’ indicates the extent to which its current plan can be adjusted to deal with the effects of the event.

For animals and agents living in social environments, an additional appraisal objective considers the implications of an event or the consequences of ones actions, in the context of social norms and moral standards. Although this objective involves many complex concepts, it only contains two global checks: *1*) ‘*external standards*’, which assesses the significance of a stimulus regarding laws, norms and moral standards common to all individuals in the same social group; *2*) while ‘*internal standards*’ estimates the extent to which an action meets personal standards, such as self-ideals and internalized moral code.

Although the brain is notorious for its massively multi-processing prowesses, in the context of the CPM, Scherer still assumes that the appraisal objectives are processed in sequence. Arguing that:

> *"This sequence assumption is justified in terms of systems economy and logical dependencies — the results of earlier SECs need to be processed before later SECs can operate successfully, that is, yield a conclusive result."* — Scherer et al. (2010, p. )

In the case of humans, and possibly other animals capable of emotions, the sequence is thus repeated endlessly. Hence, continually updating the appraisal results and giving rise to new emotional experiences. This recursive mechanism, in Scherer's opinion, accounts for the fleeting nature of emotions and is what enables animals to quickly react to unforeseen events in the environment.

#### Response patterning module

In the component process model, Scherer describes the response patterning module as including, for animals and humans at least, the Autonomic Nervous System (ANS) and the somatic system. It is, therefore, mainly involved in preparing the body for actions, motivation, reacting to events and communicating the individual's current emotional state (usually through facial expressions or changes in the voice). Compared to the appraisal module, its structure is much simpler, since every results from the SECs will simply drive the activation of a particular set of sub-components. As a result of his extensive work, Scherer was able to compile a list predicting the effects, in terms of motor and ANS activations, of the SEC results. For example, a positive result from the Novelty check, will produce an orienting response, a deceleration of the heart rate, vasomotor contraction, increased skin conductance responses, pupillary dilation, local muscle tonus changes, brows and lids up, frown, jaw drop, gaze directed, interruption of speech and action, and raising head (but see (Scherer, 2009; Scherer et al., 2010) for the complete list of predictions). The list in its current form is only intended to be used to faithfully reproduce the emotional expression of humans. Therefore, an implementation of the CPM in a virtual agent would require to adjust the different reactions to the agent's capabilities.

### Categorization module

In the CPM, a lot of importance has been given to subjective experiences, usually known as '*feelings*'. According to Scherer, as part of the categorization module, they serve a monitoring and regulation function. To fulfill its monitoring duty, the feeling component is required to integrate and synchronize patterns of changes in all other components. Scherer believes that it is this very process of integration, which gives rise to conscious feelings, stating that:

> "*If subjective experience is to serve a monitoring function, it needs to integrate and centrally represent all information about the continuous patterns of change and their coherence in all other components. Thus, feeling is an extraordinarily complex conglomerate of information from different systems. . . . The CPM suggests it is the very process of synchronization between components that elicits and organizes this process of multicomponent integration, largely outside of awareness.*"
> — Scherer et al. (2010, p.59)

However, in the context of the CPM, feelings are responsible for more than simply monitoring and reflecting changes in the different components. They have proactive roles as well, eliciting processes of cognitive reevaluation and of physiological and expressive regulation.

### Conclusion

Given the definition adopted for the concept of emotions in the CPM, a typical emotional episode can be described as beginning with a relevant stimulus entering the appraisal module. Recurring sequences of appraisal and response patterning allow the agent to react to the event and communicate its emotional states. Once all appraisal checks have reached a stable state, or conclusive results, the emotional episode ends.

Although the CPM has never been completely implemented, its appraisal module has been adopted by other research projects as part of their own models or to drive the behaviors of social robots (see Sub-Section 2.6.4 for an example of such a model, as well as Scherer et al. (2010)). Furthermore, the principles underlying the CPM are a good representation of recurrent ideas found in appraisal theories.

## 2.5 Dimensional theories

Another approach to emotion that has been considered over the years, is that of dimensional theory. Dimensional theories of emotions argue that affective phenomenon, such as emotion, mood and personality, are not discrete entities, but points in a continuous multi-dimensional space. An idea not unlike the appraisal theories describe in the previous section. Indeed, both approaches define a number of dimensions against which an object, an event, or more broadly a situation is evaluated, which leads to an emotional reaction and ultimately a change in affect. Rejecting the idea of discrete emotions, on the ground that there is little evidence for the existence of basic brain areas or circuits supporting each emotional category, dimensional theories tend to rather focus on the concept of affect. Emotions in this perspective are regarded as no more than virtual labels retrospectively attributed to particular mental or perceived body states. Therefore, the main goal implied by the adopted representation of affect is to determine the optimal number of dimensions, along with their nature, that completely describe the concept without redundancy. However, simply characterizing the position of affect at a given moment in time would be ignoring its dynamic nature. As a consequence, another line of investigation for dimensional theories involves developing a mechanism to account for changes in affect in reaction to shifts in the state of the external environment or the body. An example of such mechanism is explored introduced by J. a. Russell (2003) is explored later in this section.

Back to 1872, when Darwin noticed a similarity in the expression of emotions across species and cultures, upon closer inspection of his many portraits, at the time he drew the conclusion that there had to be a set of basic emotions that are common to all animals, and by inheritance to all humans. His discovery would later lead to the infamous theory of '*basic*' versus '*Complex*' emotions. Two years later a researcher by the name of Wundt (1874), inspired by Darwin's work on emotional expression, decided to take the mathematical route and develop a structural description of subjective feelings. Where Darwin saw evidences for a set of basic emotions, Wundt saw a potential for hidden factors that could describe emotional expressions. Through statistical analysis of introspective reports on feelings, Wundt suggested that affect could be conceptualized as a point in a three dimensional space of valence (positive versus negative feelings), arousal (calm versus exited) and tension (relaxed versus attentive). Furthermore, he believed that a mental phenomenon, described in such a space co-varies with measurable physiological changes, such as heart rate, skin conductance or visceral activity.

A century later, Mehrabian (1980, 1996) observed that, on the contrary to natural or integrated sciences like physics and chemistry, social sciences were lacking a proper set of scales (such as length, time or mass). He argued that the reason studies of affect or emotion had been handicapped for the past decades was this absence of basic dimensions, which would provide a common ground for describing and measuring emotion. Without shared values researchers could not compare their theories (a concern later shared by Scherer (2005) as well).

Mehrabian grounded his search for a set of dimensions in the expanse of evidences that other investigators had already produced. He was especially interested in the work of Osgood (1957) and noticed that the semantic differential (a method of investigation used to measure the connotative meaning of objects, events and concepts) factors of evaluation, potency and activation, suggested as a result of his experiments, were repeated consistently in studies of judgments of highly diverse stimuli. Mehrabian (1996) hypothesized that this consistency was a direct consequence of the involvement of affective reactions. Stating that "*. . . such reactions were operative in any situation.*" (Mehrabian, 1996). Mirroring the semantic differential factors introduced by Osgood, Mehrabian conceived the now famous Pleasure, Arousal and Dominance (PAD) Emotion State model, where pleasure, corresponding to Osgood's evaluation factor, denotes a positive versus negative affective state. Arousal similar to activity reflects the individual's level of mental alertness and physical activity. Finally, dominance akin to potency measures the level of control a person feels she has over her own actions and environment. Thus, a positive value for dominance is a sign that one feels in control of his surroundings and makes his own decisions without any external influence, while a negative value is synonymous with submissiveness and is characterized by a feeling of being controlled by others and/or a situation. The PAD Emotional State model was tested in relationship with 42 verbal-report scales of emotional response developed by other investigators. From the results, Mehrabian (1996, p. 263) concluded that the PAD Emotion State model "*provided a reasonably general characterization and measurements of*" a person's emotional state, since it was able to explain all the reliable variance. However, Mehrabian (1996) did not stop there and extended his Emotional State model into an Emotional Trait model. The distinction being that a trait, according to him can be computed as an average over a long period of time (months, years or even a lifetime) of a wide and representative sample of everyday affective situations. He equated an emotional trait to temperament, hence, making his scale able to measure not only the current emotional state of a person, but also her mood and on the long term one's personality. In a later paper Mehrabian (1996) described a series of experiments supporting the claim that the PAD Emotional Trait model could be used as a general framework for describing and measuring personality. It is worth noting that to this day, many scientists from all manner of fields are still using Mehrabian's distinction and definitions of emotional state and trait.

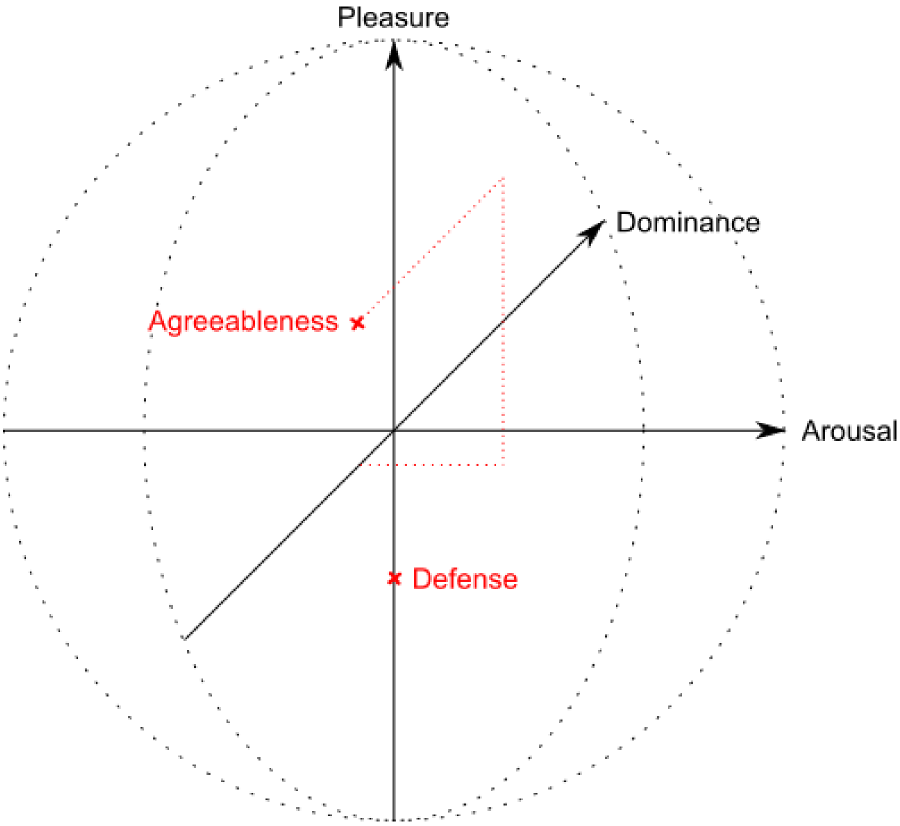


Figure 2.5 *A depiction of Mehrabian's theory of emotions. It has become somewhat of a standard model in the field of dimensional theory. In this school of thought, an emotion is represented as a point in space (exemplified here by agreeableness and defense in red). In the case of Mehrabian's model, this space is defined by three dimensions, that of pleasure, dominance and arousal. It should be noted that this three dimensional space is also used for describing personality and temperament.*

Even though, Mehrabian's Emotion State model is nowadays widely popular as a model of emotions, it bears mentioning that neither him nor Wundt (1874) gave any suggestion as to how a person's emotional state is updated as a result of her interactions with the world or changes within the body. Many researchers have since dedicated their career to fleshing out a mechanism that would account for changes in emotional state, based on alterations of the surroundings and presence of certain objects in the environment. A general consensus, which has emerged from the wealth of research performed, is that objects, events and other stimuli in a person's surroundings all have a special property called '*affective quality*'. When an object is perceived by the brain, its affective quality is interpreted and triggers a change in the person's emotional state. Effectively pushing it around, as one would a hockey puck. A concrete example of this general mechanism can be found in the core affect model developed by J. a. Russell (2003).

The framework first suggested by J. A. Russell and Barrett (1999) and further detailed by J. a. Russell (2003), is a direct result of the observation that psychological theories related to emotions rarely define the concept and usually rely on emotion words such as fear, anger or joy. In Russell's point of view, those emotion words are nothing more than folk concepts, used long ago by our hunter-gatherer ancestors. Since these concepts are not supported by any scientific theory, they have no place in psychology, or any other field for that matter. Therefore, Russell sought to establish a framework describing the structure underlying emotion.

To create such a framework J. a. Russell (2003) needed to define the primitive building blocks, which through different combinations account for all the observations and manifestations of emotions. In this endeavor, Russell followed in Oatley and Johnson-laird's (1987) footsteps. They argue that emotions supported by cognitive processes, are always directed at an object, whether real or imaginary. Therefore, those emotions are not primitive elements, but complex events. Instead they suggest that affect should be based on emotional processes that can exist free of any external objects. That is not to say that external objects or events do not influence emotions in any way, but rather that the core structure of emotion should not rely on external stimuli for its definition.

As discussed in Section 2.5, dimensional theories of emotions provide an empirically established solution to this problem. Experiments conducted within this field have gathered much evidences pointing to the dimensions of pleasure — displeasure (valence) and activation — deactivation (arousal) as being primitive and ubiquitous. As a consequence, Russell adopted a combination of these two dimensions, which he refers to as '*core affect*', as the first primitive of his framework. The perception of the pleasant and arousing qualities of external stimuli, he defined as the second primitive, and labeled it '*perception of affective quality*'. The term perception is intentionally used in this context, since the perception of a stimulus' affective quality is described as being indistinguishable from the cognitive perception of its other properties.

Core affect then exists within the individual and can be conceptualized as a free floating point in the two dimensional space of pleasure and arousal. Whereas, affective quality, which can be redefined as the propensity to alter core affect, is external and characteristic of each stimulus. Considering that core affect is an object-less representation of the individual's current emotional state, a mechanism is required to link a change in core affect to a perceived cause. To fulfill this role, Russell introduces the '*attributed affect*', which is defined by three necessary and sufficient features: *1*) a change in core affect; *2*) an external object perceived as the cause for the change; *3*) and the process of attribution of the core affect to the object. Even though, attributed affect is a cognitive process, it is not perfect and has been proved to give rise to misattribution (J. a. Russell, 2003). In misattribution, a change in core affect, caused by one source, is attributed to another target, which leads the individual to perceive the target as more pleasant, or less so, than it really is.

Russell argues that combined with information processing and behavioral planning, the adopted primitives account for the myriad manifestations and influences called emotional. For one, in the same way other dimensional theories do, core affect qualifies in some particular cases as an emotion. For example, a simple core affect of pleasure, could be labeled as happiness, and core affect of displeasure and high arousal may qualify as anxiety. Nevertheless, pleasure and arousal alone are not enough to account for most emotional episodes. Russell advocates for a perspective integrating the dimensional approach with the categorization process found in most appraisal theories.

Over the course of its existence dimensional theories have received a lot of criticisms. Chief among them is the fact that, similarly to early theories of emotions, the emotional episode considered has no clear beginning. In other words, dimensional theories usually lack a mechanism to discriminate between objects that should trigger a reaction, and any other object that can safely be ignored. Furthermore, since the main focus of the field is to find a suitable multi-dimensional space to describe and measure an individual's emotional state and personality (emotional trait), very little attention has been given to the relations that exist between emotions, and even less so to the many interactions the affective and cognitive systems benefit from, in the brain.

One famous exception can be found in the work of Plutchik (1958, 1962); Plutchik and Conte (1997), which culminated in the well-known '*wheel of emotions*'. Inspired by the parallel between emotions and colors, which McDougall (1921) depicted as follow:

> *"The color sensations present, like the emotions, an indefinitely great variety of qualities shading into one another by imperceptible gradations; but this fact does not prevent us regarding all these many delicate varieties as reducible by*

> *analysis to a few simple primary qualities from which they are formed by fusion, or blending, in all proportions . . . the same is true of the emotions.*" — McDougall (1921, p. 114)

Plutchik (1958) sought to decipher and model the relations between emotions. In 1958, he introduced a cone-shaped model based on the dimensions of pleasantness — unpleasantness, attention — rejection and intensity (Plutchik, 1958). At the same time he also suggested that there are eight bipolar basic emotions: joy versus sorrow, anger versus fear, acceptance versus disgust and surprise versus expectancy.

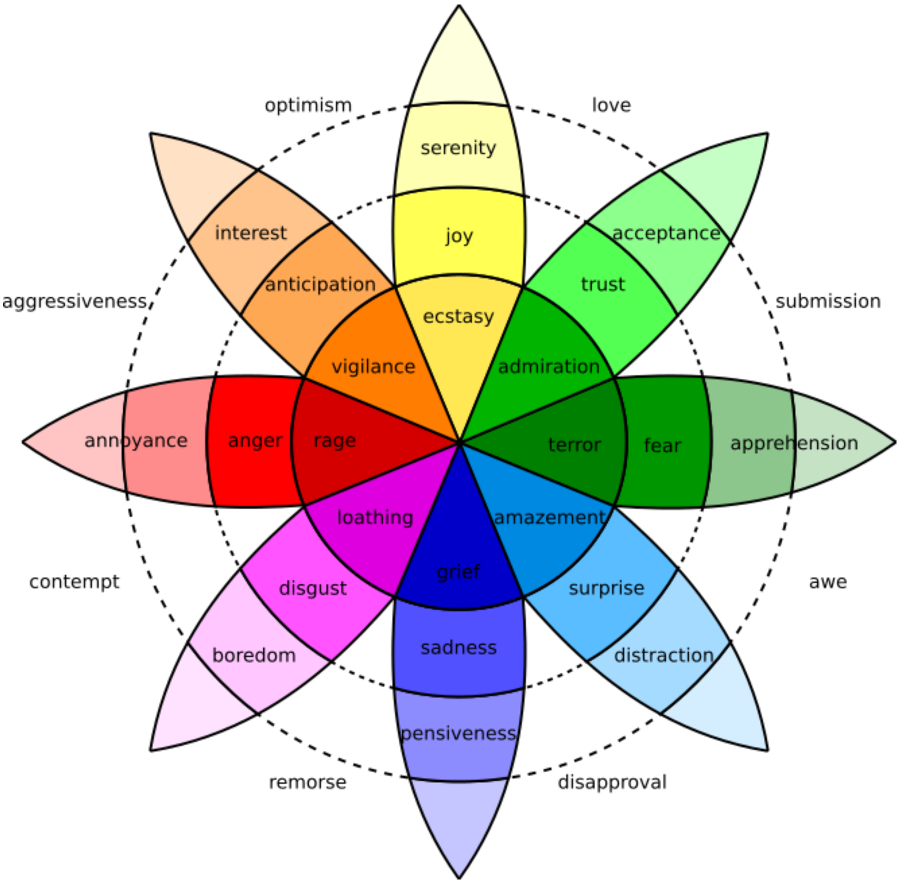


Figure 2.6 *A well-known representation of Plutchik's mechanism for creating higher-level emotions, in which they are compared to colors. As is the case with colors, a select few are defined as primary. Blending two primary colors together, results in a new color. In the same way mixing two primary emotions, gives birth to what Plutchik refers to as a dyad. A first level dyad is a blend between two adjacent emotions. The level of dyad increase the further apart the primary compounds are.*
Source: commons.wikimedia.org.

In a later study Plutchik (1962) revised his model from a cone to a circle (see Figure 2.6). For this experiment he used a modified pair-comparison method, in which a set of three non-synonymous emotions were chosen. Participants then had to rate, on a bipolar scale, the similarity of 146 emotion words to each of the three reference words. The scale used for rating was comprised of 11 points ranging from *opposite* ($-5$), through *no relation* (0), to *the same* (5). The mean similarity ratings were then converted to angular position on a circle. Emotion words with opposite meaning sitting diametrically, while emotion words with no relation were separated from each others by a 90 degrees angle. After, factor analyzing the results of the experiments, Plutchik concluded that two factors accounted for most of the variance, hence giving birth to the aforementioned wheel of emotions. To further prove the validity of this emotion circle, a second experiment was conducted using the independent semantic differential method, introduced earlier by Osgood (1957). A factor analysis of the results revealed that the circles from both experiments had a product-moment correlation of 0.90, indicating that they were almost identical. Hence, lending further credit to Plutchik's (1962; 1997) '*Circumplex model of emotions and personality*'.

Recently, with the advent of service robots and the development of the game industry, dimensional theories are experiencing a resurgence in popularity. Considering that these theories allow for the reduction of complex concepts, such as affect, mood or personality, to simple vectors, as well as providing a mechanism to account for the influence of external object or event on the emotional state of any given person, it is easy to understand why people outside the field of affective science would be attracted to dimensional theories. They provide a mathematical representation convenient for further processing or interpretation whether by a computer program or a human. Furthermore, recent studies have shown that when used in artificial agents, they performed better in emotion expression and recognition tasks (Marsella et al., 2010, p. 29), then when using appraisal theories. Although, only semantic models have been presented in this section other theories and models subscribing to the dimensional approach to emotions can be found in Davidson (2003); Davidson, Scherer, and Goldsmith (2003), as well as Fontaine, Scherer, and Soriano (2013) to name but a few examples.

## 2.6 Anatomic theories

The dawn of the nineteenth century, brought with it a peculiar movement later known as '*phrenology*' (Damasio, 2008; LeDoux, 1998). Phrenologists were scientists, or rather pseudo-scientists by current standards, who studied personality traits and mental disorders by feeling

Figure 2.7 *An example of a map used by phrenologists to locate human faculties on the skull.* Source: commons.wikimedia.org. Author: Mirokado.

the topography of a person's skull. In 1796, Franz Joseph Gall, a reputed scientist at the time, began to give lecture on '*organology*', which can be best described as a mix of early neuroscience, psychology and applied philosophy. Through organology, Gall introduced radical concepts from which phrenology would stem, some of which are now widely accepted, while other had to be abandoned. Contrary to the dualist belief spread by psychology so far, Gall suggested that the brain as an organ is the seat of the mind. He went on in his reflection and introduced the controversial idea that the brain was not in fact one unique organ that produced the mind and its different faculties, but rather an aggregate of many organs each with its own psychological faculty, and each working independently from one another. The many organs that made up the brain, in Gall's theory, were called '*centers*'. Thus, there was a center for sensation, one for feeling, another for memory and language, and some more exotic ones for veneration, benevolence, friendship, sublimity, suavity and philoprocentivness (what ever that might be) to name but a few. Where Gall's doctrine took a turn for the worse, was with his belief that the '*power*' or mental faculties of each center was proportional to its size. Furthermore, he argued that a person's skull would have small bumps over those areas which were well developed, while smaller centers would result in dips in the shape of one's head. Therefore, it was possible to derive someone's personality by simply '*reading*' the surface topography of his head (an example of a map drawn by phrenologists and depicting the location of the different personality traits can be seen in Figure 2.7).

Being born in the middle of the Age of Enlightenment, an era where science was fashionable, phrenology grew in popularity and quickly spread throughout Europe. Soon phrenologists everywhere were publishing maps describing the locations of the different faculties on the human skull. By 1832 it even reached the United States, where an accident of great importance for affective neuroscience was about to happen.

The year is 1848, Phineas Gage is a construction foreman working for the Rutland & Burlington Railroad company. At the head of a large group of men, also called a '*gang*', their job is to lay down new tracks for the railroad's extension across Vermont. Rather than twist and turn around each outcrop of hard rock, the strategy is to blast a straighter and level path through them. A process that requires to drill holes in the rock and pour in explosive powder. Sand is then placed on top of the powder and '*tamped in*' to contain the explosion and direct it inside the rock. Phineas was usually in charge of pounding the sand with his iron rode. However, this time acting while distracted Gage pounds directly on the explosive powder, igniting it. The explosion, unrestrained by the sand, projects the iron rode through Gage's skull with enough force for it to land some twenty meters away (Damasio, 2008). In the hands of Harlow (Damasio, 2008; J. Harlow, 1848; J. M. Harlow, 1868), Phineas Gage would make a miraculous recovery in less than two months. However, as his friends, coworkers and family would soon discover: "*Gage was no longer Gage*" (J. Harlow, 1848). For the duration of Gage's recovery and the next twenty years of his life, Harlow recorded in great details the ways in which Gage's personality had changed. Although, Phineas had regained his full range of motion, his dexterity and could see as well as speak without any difficulty, his temperament was different. Before his accident Phineas had been a respectable person with a bright future ahead of him. Whereas after, J. M. Harlow (1868) would describe him as:

> "*. . . fitful, irreverent, indulging at times in the grossest profanity which was not previously his custom, manifesting but little deference for his fellows, impatient of restrain or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operation, which are no sooner arranged than they are abandoned in turn for others appearing more feasible. A child in his intellectual capacity and manifestations, he has the animal passions of a strong man.*" — J. M. Harlow (1868, p. 13–14)

While Phineas Gage died on the 21$^{st}$ of May 1868 in San Fransisco, his case through the thorough report of Harlow, would provide irrevocable proof for one of phrenology's tenets (Damasio, 2008). Since only a small portion of the brain was lesioned, triggering a change in personality, while all other aspects of Gage's capabilities remained intact, the conclusion drawn by early 1920s neuroscientists was that the brain is indeed an aggregate. However, instead of being called centers the brain's components are nowadays called systems each fulfilling a precise function, while influencing each others. Furthermore, Gage's change in temperament hinted at the fact that within the brain the mind shares its seat with emotions.

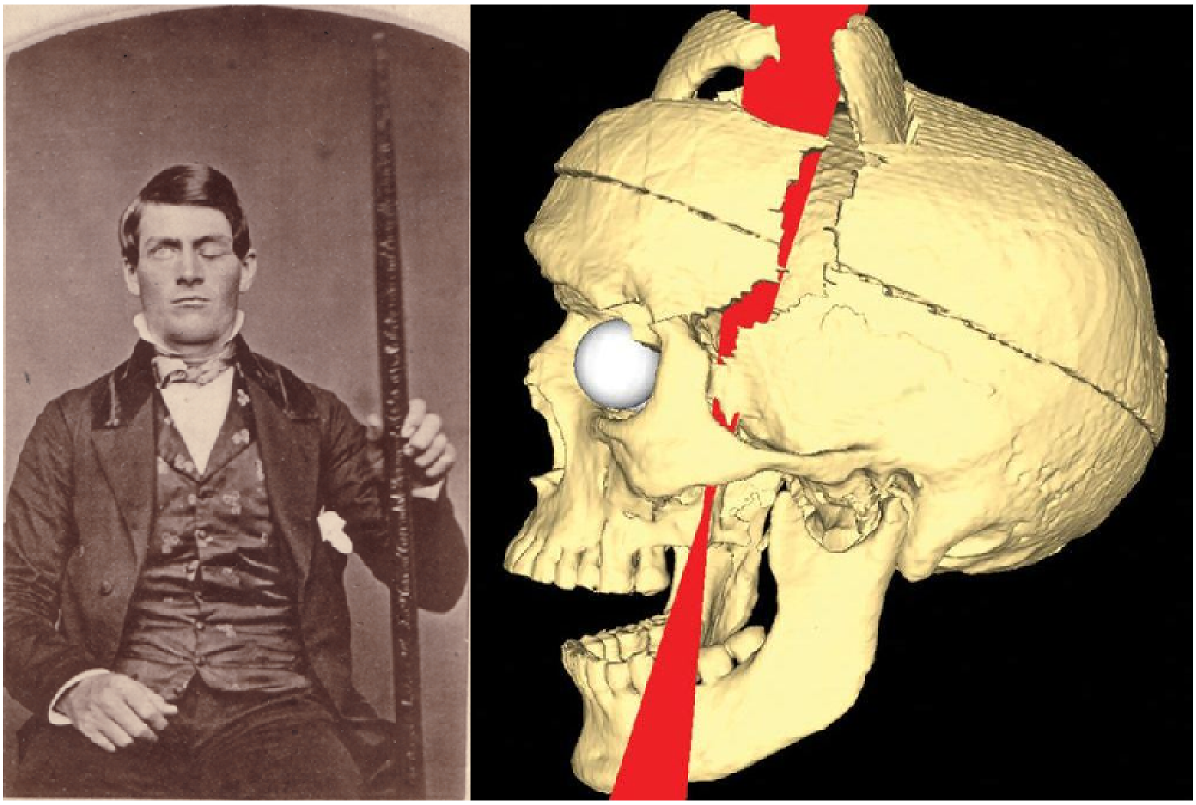


Figure 2.8 *On the left: A picture of Phineas gage holding his tamping iron. On the right: A virtual simulation of the trajectory the iron rod followed through gage's skull.*
Source: commons.wikimedia.org

This idea of a brain made of systems and being a container for both the mind and emotions, along with theories that Darwin (1872) would put in writing only half a decade later marks the most important turning point in the history of affective science. As a matter of fact, in Section 2.2 this point in time has been defined as the birth of affective science. This nascent field saw the quick rise in popularity of the James-Lange theory (see Section 2.2 and James

(1884) for more details) followed by a long period of '*stagnation*'. Until Cannon et al. (1927) armed with a better understanding of the different areas that make up the brain, as well as the connections between them, broke through the silence to suggest an model centered around the interaction between the thalamus and the cortex (a described in Section 2.3).

Exactly ten years later, upon learning that an American benefactor had provided a generous financial support to a British laboratory, whose project consisted in figuring out how emotions work, Papez (1937) his national pride hurt published his famous theory in a few days, or so the story goes. Even though, the goal at the time was to show that Americans had some ideas about emotions too, Papez (1937) had just introduced one of the most influential model of emotions in the brain. Based on the work of Cannon et al. (1927), Papez (1937) used the following definition of emotion as a basis for his reflection:

> *"The term "emotion" as commonly used implies two conditions: a way of acting and a way of feeling. The former is designated as emotional expression; the latter, as emotional experience or subjective feeling. The experiments of Bard (1929) have demonstrated that emotional expression depends on the integrative action of the hypothalamus rather than on that of the dorsal thalamus or cortex, . . . For subjective emotional experience, however, the participation of the cortex is essential. Emotional expression and emotional experience may in the human subject be dissociated phenomena."* — Papez (1937, p.726)

Consequently, Papez's theory describes the emotion system as a circuit, hence, its later labeling as the '*Papez circuit*', in which sensory inputs into the brain are split, at the level of the way stations in the thalamus, into the '*stream of though*' and the '*stream of feeling*' (Dalgleish, 2004; LeDoux, 1998; Papez, 1937). The stream of thought depicts a path between the thalamus and the lateral areas of the neocortex, where sensations are turned into perceptions, thoughts and memories. The stream of feeling channels sensory data, coming into the brain via the thalamus, directly to the hypothalamus, or more precisely the hypothalamic mammillary bodies, which generates emotions. The outcome of hypothalamic processing is transfered to the anterior thalamic nucleus and is relayed to the cingulate. The cingulate cortex receives activations from both the stream of thought and stream of feeling. It was described by Papez as the place where "*environmental events are endowed with an emotional consciousness*". Finally, the cingulate cortex projects to the hippocampus, which in turn completes the circuits by way of connections with the hypothalamus, regulating the brain's emotional response. Through this circuit, Papez suggested that emotional experience could be generated in two different ways: by simple activation of the stream of feeling by sensory objects; or as a consequence of information flowing through the stream of thought, where the stimuli would be perceived and memories activated. Reaching the cingulate cortex, perceptions and memories would be interpreted and give rise to feelings.

Around the time Papez wrote his theory, the connections described were poorly understood or even unknown, because the methods for tracing such synaptic projections were crude. Therefore, based on the known clinical effects of brain damage to the different regions involved in the circuit, Papez speculated their existence and function within his circuit. Especially for the hippocampus and the cingulate cortex, were lesions results in intense emotional, convulsive and paralytic symptoms: delirium, depression, loss of emotional spontaneity and sometimes coma, respectively.

Although, all of the connections hypothesized by Papez (1937) have since been found to exist, there is not has much evidence showing that the different areas suggested to be part of the circuit are indeed central to the production of emotions in the brain (Dalgleish, 2004).

The same year that saw the publication of the Papez circuit, witnessed the report of an accidental discovery, by Kluver and Bucy (1937), of the effects of damage to the temporal lobe in monkeys. In their study, they consistently observed that when the temporal cortex is lesioned, animals (monkeys in this case) as well as humans experience '*psychic blindness*'. Meaning that the visual acuity of the subject is not degraded in any way, but the animal is incapable of grasping the psychological significance of a stimulus. For example, monkeys presented with different objects, such as a comb, a sunflower seed, a screw or a live snake, will inevitably bring each object to their mouth before deciding which is edible. Other characteristics of the '*Klüver-Bucy syndrome*' include hyper-sexual activity, where animal will try to copulate with other members of the same sex or with members of other species and a certain tameness in the presence of previously feared objects.

Building on the work of Papez (1937), Klüver and Bucy (1937), and extending it with Freudian psychology, MacLean (1949) sought to establish an all-encompassing theory of the emotional brain, similar to what Arnold would do for the field of appraisal theory more than a decade later. MacLean was also the first, since Darwin (1872), to really consider the brain along its fourth dimension: time. As is detailed bellow, his theories not only describe the different brain structures and their involvement in emotions, but also examine the brain's evolution and its implication for emotions. Aware of the attention received by both the cortex and hypothalamus, in the work of his predecessors, for their role in emotional experience and emotional expression respectively, Maclean focused on uncovering the ways these two brain areas communicate. Believing this to be the key to understanding how the affective quality of an experience could influence the autonomic nervous system and generate an emotional reaction, as well as lead in the most extreme cases to psychosomatic diseases like hypertension. MacLean, like Papez before him, was convinced that for an animal to be capable of experiencing and discriminating between emotions, it needed a cerebral cortex. At the time, though, no significant connections between the newly evolved neocortex and the hypothalamus were known, making it impossible for the former to exert any kind of influence over the emotional reaction process. The solution was to be found in Papez's circuit, which established that the evolutionarily older medial cortex, called the '*rhinencephalon*', has strong ties with the hypothalamus. Further, noting that stimulation of the rhinencephalon, but not areas of the neocortex, produced autonomic responses, MacLean renamed the medial cortex into the '*visceral brain*'. He also concluded that the visceral brain had to be the seat of emotions in the brain.

According to MacLean, because the neocortex had not evolved yet, primitive animals were limited to the visceral brain as the highest center for coordinating behavior. In these creatures, the visceral brain was in charge of all the instinctual behaviors and basic drives underlying the survival of the individual and to a greater extent of the specie. However, the advent of the neocortex in mammals opened the door for higher forms of psychological functions. Therefore, the neocortex incrementally took hold of the body's musculature, all the while fulfilling the higher functions of intellect. Yet even in humans, considered to be at the pinnacle of brain evolution, the visceral brain remains essentially unchanged and fulfills its survival duty.

Drawing from his previous observations, MacLean hypothesized that emotional expression resulted from the integration of sensations arising from the external environment with visceral sensations. He suggested this integration to take place inside the visceral brain, where stimuli from within and without the body were transmitted as nervous impulses to '*cerebral analyzers*'. Similar to Papez, MacLean designated the hippocampus as the brain area containing the cerebral analyzers. As a consequence the hippocampus became the main component underlying emotions. Moreover, acknowledging the limitations of the hippocampus as a more primitive structure, compared to the neocortex, he argued in a fashion reminiscent of the Cannon-Bard theory, that the hippocampus needs the top-down control of the neocortex to keep emotional reactions in check. Without inhibition from the neocortex, according to MacLean, the hippocampus would run wild, leading to phobias or obsessive-compulsive disorders.

After refining his hypothesis for three years, MacLean (1952) introduced the concept of the '*limbic system*' as a replacement to the visceral brain. In addition to the brain areas already present in the Papez circuit, the limbic system included such regions as the amygdala, the septum and the pre-frontal cortex. Extending the visceral brain hypothesis, the structures of the limbic system are primitive components working together to ensure the survival of the individual and the specie. This system evolved to mediate visceral functions and affective behaviors, such as feeding, defense, fighting and reproduction. It is therefore, an extension of the visceral brain as it not only maintains homeostasis, but also underlies the emotional life of the individual.

Finally, almost twenty years later, MacLean (1970) published the '*triune brain*' theory, according to which the forebrain has gone through three stages of evolutions, namely: reptilian, paleomammalian and neomammalian. The paleomammalian stage roughly corresponding to the aforementioned limbic system. As a result, the human brain can be conceived as a three layered hierarchy. Where each layer has its own intelligence, memory, sense of space and time, motor control and other specialized functions. In MacLean's perspective, the brains of humans, primates and advanced mammals, all have three layers. Lower mammals have the reptilian and paleomammalian layers, while other vertebrate creatures, such as birds, reptiles, amphibians and fishes, are left with only the reptilian layer. The triune brain is MacLean's most global theory and offers a broad view of the brain, and the limbic system's place within this ensemble, through the scope of evolution. At the same time it provides a flexible architecture to explain the differences in levels of complexity between the capabilities, mental as well as behavioral, of advanced mammals in comparison to lower mammals and vertebrates.

In consequence of the depth and richness of MacLean's work, the surface of which has only been scratched in this chapter, it is easy to see why by 1952 most affective neuroscientists believed the search for the seat of emotions in the brain to be over. It also serves to explain why the concept of the limbic system has survived through history and is still a theory taught to young neuroscientists today. Be that as it may, the survival of the limbic system was not without problems, as it has received its fare share of criticisms. Some scientists went as far as arguing that the concept should be abandoned all together. The reason behind this argument being that in the context of our current understanding of the brain, the limbic system does not have well defined borders anymore. When MacLean first introduced the concept, he used the phylogenetic properties of its component to identify its borders, as mentioned above. Shortly after the publication of MacLean's theory, researchers sought to confirm his reasoning and found that indeed vertebrates seem to be lacking a structure resembling the neocortex, while still having a medial cortex. Hence, lending more support to the limbic system concept and MacLean's triune brain theory. Yet in the early 1970s, anatomists like Karten and Northcutt (Karten & Shimizu, 1991; Northcutt & Kaas, 1995) were able to show that, even though primitive animals did not have a neocortex proper, other regions of their brain meet the structural and functional criteria of the neocortex. As a result, it was no longer possible to distinguish brain areas based on their emergence in evolution, rendering the whole rhinencephalon, visceral brain and limbic system suspect. Many of MacLean's followers have since tried to revive the concept by providing other criterion to find the regions included, but to no avail. Even though the concept of the limbic system might soon be abandoned, it does not mean that all of MacLean's ideas were misguided. His approach to emotions through the lens of evolution and his focus on the capabilities the emotion system affords animals, still serves as guide in current research projects.

### 2.6.1 The rise and fall of the '*basic emotions*' theory

Since Darwin speculated the existence of innate emotions, which are shared not only across cultures, but probably with most mammals and even close related vertebrates as well, many have tried to identify those so called '*basic emotions*' and the neural circuitry underlying them. In Darwin's mind the universality of a particular emotion was characterized by the similarity of its expression across different civilizations. Further, in his perspective basic emotions represented innate, patterned responses that are controlled by '*hardwired*' brain circuits. His claims relied on simple portraits that he had drawn himself, while traveling around the world. To lend more credibility to Darwin's presumptions, modern researchers have gone into remote areas of the world to establish through rigorous scientific methods that at least a subset of the emotional range humans are capable of, have a universal mode of expression. As a result, the first list of basic emotions proposed (LeDoux, 1998; Scherer et al., 2010), included: surprise, interest, joy, rage, fear, disgust, shame and anguish. On the one hand, relying on facial expression alone Ekman (1973, 1992) was able to produce an overlapping list including: surprise, happiness, anger, fear, disgust and sadness, which is still regarded as a standard in most current research projects on emotions recognition and synthesis in virtual agents or robots. On the other hand, scientists like Plutchik (1958, 1962) and Frijda (1986) argue for a different approach. Considering that, the further down the evolutionary scale one goes, the less the creatures are capable of facial expression. Yet the wealth of emotional expressions, observed in mammals, is still available through body language. From his observations, Plutchik deduced yet another list of basic emotions that largely overlaps with Ekman's one, with the addition of acceptance and anticipation. The last approach, considered has been suggested by Panksepp, Mir, Delgado, and Edwards (1982), who used the behavioral consequences of electrical stimulation of areas of the rat brain to reveal four basic emotional response patterns: panic, rage, expectancy and fear.

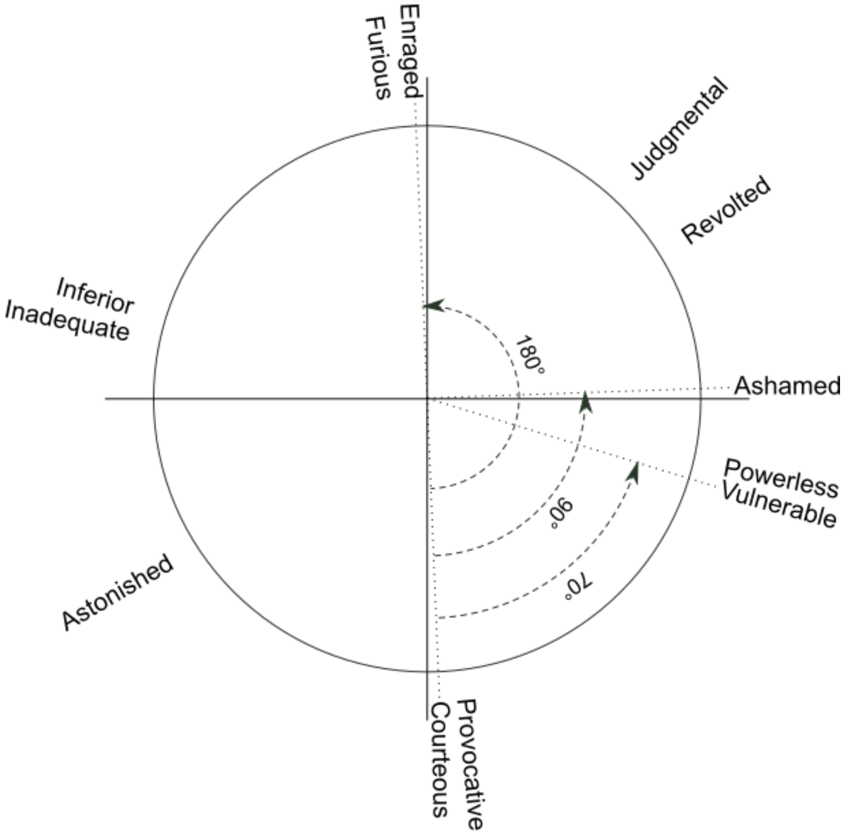


Figure 2.9 *This figure presents a graphical depiction of Plutchik's circumplex model of emotions and personality. To establish the validity of his model, he performed a survey in which participants had to rate the similarity of* 146 *emotions. To visually represent the results of his experiment, he chose to use a circle, on which emotions with opposite meanings sit diametrically opposite from each other. On the other hand, two emotions separated by a* 90 *degrees angle have no relation at all. Therefore, the dissimilarity between two emotions increase with the angle between them. As was the case for Mehrabian before him, Plutchik's model also describes both emotions and personality.*

Far from denying the existence of any other non-basic emotions, most theorists conceive of them as the results of blending or mixing of two or more basic emotions together. Among the mechanisms suggested to account for this blending process, Plutchik's theory of the circumplex (Plutchik, 1962; Plutchik & Conte, 1997) contains the most sophisticated one. As mentioned in Section 2.5, in the circumplex theory, emotions are placed on a circle with the angular position of one emotion relative to another being indicative of their relation (as illustrated by Figure 2.9). Thus, 180 degrees implies opposite emotions, while 90 degrees means that the two emotions have no particular relation. Finally, emotions separated by only a few degrees have a very close meaning. However, Plutchik's theory goes beyond this model and gives a mechanism to account for non-basic emotions. A complex emotion, similar to how colors work, is a blend between two basic emotions on the circle, hence the name '*dyad*'. A blend of two adjacent basic emotions is called a '*first order dyad*'. Blends between two basic emotions, which are separated by one other emotion, are referred to as second order dyads, and so on for higher order dyads. A blend between two nearly opposite emotions (high order dyad) can result in conflict, which in itself can be the cause for a new emotion. For example, joy and fear are separated by acceptance and their fusion is imperfect. The resulting conflict is the source of the emotion called guilt.

In Plutchik's perspective the blending mechanism is typically a cognitive operation. Hence, basic emotions are shared across species, but non-basic emotions, being the result of cognition, can only be shared by animals with the same cognitive capabilities. As a result, most if not all non-basic emotions are uniquely human.

Although the basic emotion theory has enjoyed an era of popularity, it has not been without a cost. Throughout history, it has also received its fair share of criticisms. Some appraisal theorists considering emotions to be the internal representation and interpretation (appraisal) of situations, argue that emotions are social constructions, and not the result of mindless processing of biological hardware. Therefore, in this so called '*social constructivist*' perspective, emotions are purely inventions of the human society, resulting from our need to categorize or interpret the world around us, as opposed to biological phenomenon. As a consequence, the concept of innate basic emotions cannot exist in such framework.

Ortony and Turner (1990) published a paper, in which they questioned the foundations upon which all basic emotion theories are built. A recurrent complaint in the field of affective science, even to this day, is the lack of agreed upon definition for the very concept of emotion. While, this is true for the whole field, the problem is compounded when it comes to basic emotion theories. Not only does the concept of emotion change according to the author's point of view, there is also no consensus as to what characteristics a given mental state should possess to be considered a basic emotion, a fact that Ortony and Turner target with three of their complaints. First, as is made clear by the different lists of basic emotions detailed before, there usually is some overlap from one list to another, but from author to author the size of the list will change, and so do the label attributed to each basic emotional state (the lists given above being standardized to make the overlap clearer). The shorter lists counting only two emotions (pleasure and pain for example), while the longest ones can contain up to 18 emotions.

Further, some of the lists contain mental states, that Ortony and Turner (1990) argue are not emotions. The prevalent examples are interest or surprise. Assuming that a given state should at the very least require a non-neutral valence to constitute an emotion. In such a case interest does not meet the requirement for being an emotion, since the state of interest in itself does not have any valence. The valence is attributed to the object target of the attention mechanism. Interest can be a cause for the appearance of an emotion, but is not one. The same holds true for surprise. The third complaint expressed by Ortony and Turner (1990) concerns the wealth of methods used in studies related to basic emotions. A direct consequence of the lack of established characterization, is that each research project employs its own experimental methods to determine a set of basic emotions, ranging from introspection, analysis of the words describing emotions, to brain stimulation. This again results in non-overlapping, sometimes very different, lists of basic emotions.

The final criticism is directed toward a smaller group of basic emotion theorists. The point of view adopted by the different members of this group, is the result of an alternative interpretation of Darwin (1872)'s theory. As mentioned earlier, using the principle of the '*survival of the fittest*' Darwin (1872) was able to deduce that emotions were a mechanism that evolved from a common ancestor. As a result, some emotions are shared not only by all humans, but also throughout the animal kingdom. Furthermore, it can be deduced that the fundamental brain structures involved in those emotions are also conserved across species. The role of those conserved circuits, however, is rather vague. The adopted consensus among basic emotion theorists, thought, is that a hardwired neural structure underlies each basic emotion. These neural populations are thought to be in charge of producing the emotion, as well as the accompanying innate reaction (Ekman, 1973). Yet so far there has been little evidence to support this idea. Rather the different clues seem to indicate the existence of a global emotion or '*response*' system. Indeed, evidences for the existence of hardwired structures supporting reaction patterns associated with emotions have been discovered. However, the link between these response patterns and the different emotions observed is one of correlation, but not one of causation, as the basic emotions theorists interpreted it (Ortony & Turner, 1990).

The consequences of Ortony and Turner's publication was a decrease in popularity for the theory of basic emotions. However, the idea has managed to survive and still strives to this day, where it inspires the work of some scientists (Damasio, 2008; Frijda & Parrott, 2011) or sometimes only appears implicitly (Wager et al., 2015). While others do not hold it in high regards anymore or even outright reject it. Part of its continuing success can also be attributed to its use in machine learning especially deep-learning projects where lists of basic emotions provide a convenient foundation for categorization.

### 2.6.2 The '*Survival Circuit*' theory

In spite of its declining reputation and many detractors, the principles underlying the basic emotions theory are worth reinterpreting in different directions. The core idea of circuits that evolved to fulfill specific tasks and are inherited from our animal ancestors has recently been revisited. Instead of looking for mechanisms producing specific emotions, LeDoux (2012) suggests to focus on the functions fulfilled by emotions and identify the underlying brain circuits. This idea is by no means original, since it echoes a similar perspective adopted by Sloman (2001) a decade earlier. On the contrary to Sloman, however, LeDoux introduces a concrete implementation of this idea in the shape of the '*survival circuits*' theory:

> "*By focusing on survival functions instantiated in conserved circuits, key phenomena relevant to emotions and feelings are discussed with the natural direction of brain evolution in mind (by asking to what extent are functions and circuits that are present in other mammals also present in humans) rather than by looking backward, and anthropomorphically, into evolutionary history (by asking whether human emotions/feelings have counterparts in other animals).*" — LeDoux (2012, p. 2)

Furthermore, by focusing on these preserved circuits, the theory goes beyond simply explaining the mechanisms underlying emotions. It offers a unified framework in which related phenomena, such as motivation, reinforcement and arousal, are considered basic components.

At its core, the survival circuit theory exploits the relationship between innate survival functions and emotions, which was first proposed by Darwin (1859, 1872), to explain the emergence of emotions from the interaction of multiple brain systems. It is worth mentioning that, the idea of innate survival functions, inherited from our animal ancestors, originated the consensus which many neuroscientists take for granted, that the neural circuits at the base of the brain are innately hard wired by evolution and maintain the well-being of the organism (an idea explored in Section 2.6 and again in Sub-Section 2.6.3). In turn, this idea led to the theory of basic (hard-wired/innate) emotions, which is nowadays a major controversy in the field of affective science.

The survival circuits, as described in LeDoux's theory, are responsible for maintaining the organism's homeostasis: maintenance of energy and food supplies, thermoregulation, defense and reproduction. Through their life-sustaining and adaptive functions, therefore, they increase an animal's chances of survival and reproduction, in complex, dynamic and sometimes hazardous environments. Further, LeDoux argues that, although the adaptive responses to a given threatening event is species-specific, much of the organization for the structures which make up the survival circuits have been conserved. According to him, this is evidenced by the fact that even single cell organisms are capable of detecting and avoiding harmful chemicals, while moving toward substances rich in nutrients. Through evolution and pressured by the environment, animals have grown in complexity and, therefore, in capacities to implement elaborate adaptive strategies. However throughout the animal kingdom, and even in humans, the mechanisms underlying each survival function have been conserved.

On the one hand, the survival circuits are quite similar to the innate circuits, that basic emotions theories hypothesize account for those shared and universally recognized emotions. Indeed, both approaches rely on the existence of inherited neural populations nested at the base of the brain. On the other hand, while basic emotion theories use those brain systems to explain the origin of emotional experiences, LeDoux's (2012) theory focuses on the survival functions fulfilled, but do not make any claim about these systems originating any feelings. As described later, the survival circuits do play a part in the emergence of emotional phenomena, however they are not considered to be the origin, only the trigger. As a consequence, within the survival circuit theory, there are no circuits dedicated to anger or happiness, for example, however there are systems accounting for defense, thermoregulation, maintenance of energy and food supplies, fluid balance, and reproduction. Each system, in turn, interacts with basic control components to fulfill its role, depending on the organism's motor and sensory capabilities. This last step being where species-specific adaptive responses are finally implemented.

Finally, each survival circuit is characterized by a set of detectors, implemented in neural populations, which monitor the stream of input sensory data in search for stimuli relevant to the circuit's function. Similarly to what neuroscientists usually assume, LeDoux argues that the configuration of the neural population, underlying each detector and survival circuit as a whole, has been inherited from our animal ancestors. This allows the circuits to detect, and react to, specific stimulus patterns using tried and true strategies refined through natural selection. Additionally, survival circuits can also learn to respond to novel stimuli, through its association with an innately triggering event, a process usually referred to as '*classical conditioning*' (see Sub-Section 3.2.1 for a detailed definition).

In summary:

> "*survival circuits are sensory-motor integrative devices that serve specific adaptive purposes. They are tuned to detect information relevant to particular kinds of environmental challenges and opportunities, and they use this information to control behavioral responses and internal physiological adjustment that help bring closure to the situation. All complex animals (invertebrates and vertebrates) have survival circuits. Core components of these circuits are highly conserved in vertebrates.*" — LeDoux (2012, p. 5)

The brain is an integrated system made of multiple sub-systems that interact to meet the challenges and opportunities afforded by the environment. This holds true for the survival circuits, which interact with each other via inhibitory connections, so that only one response is enacted. For example, in the presence of a threat, any reproductive or eating behavior is inhibited, in favor of fleeing or fighting, depending on the context. The inhibition or activation of a given circuit depends on its overall activation value relative to the values of the other circuits. Hence, in the previous example, even if at first an animal might flee in the presence of danger at some point it will be compelled to face the threat to gather food supplies, since fleeing uses precious metabolic resources needed for any of the survival functions.

Innate and learned emotional stimuli within the survival circuit theory are equivalent to conditioned and unconditioned circuit triggers. LeDoux argues that in addition, they can also be described as incentives and reinforcers. Where an incentive is taken to be a stimulus that motivates instrumental behavior, and a reinforcer is a stimulus that increases the probability of its associated instrumental behavior to be learned and later performed in response to a given context. The multiple roles assumed by the same stimulus are made most apparent in an experimental paradigm called instrumental conditioning (explored in more details in Sub-Section 3.2.1). A standard instrumental conditioning experiment usually involves an animal, that has to perform a certain action, such as pressing a button or pulling a lever, for example. Upon completion of the instrumental behavior, the animal receives a reward in the form of food pellets, orange juice or any other culinary delicacy. The stimulus of interest in this case is the reward, which is designed to trigger one of the innate circuits. In addition to triggering a survival circuit, this stimulus also serves to inform the animal that performing this last action was a correct reaction, given the context. Therefore, it is a reinforcing signal as well, increasing the probability of said action to be executed should the situation present itself again. Finally, once learning is achieved, a stimulus signaling a deficiency in energy or food supplies, will trigger the associated survival circuit. Furthermore, in the case of instrumental conditioning, since no source of energy is immediately available, the same stimulus also acts as an incentive for the animal to press the button or pull the lever to get the reward.

A major consequence of innate and learned stimuli assuming those multiple roles is that emotion related concepts, such as motivation and reinforcement, can be explained in terms of survival circuit as well. This avoids having to resort to aspects of emotions to describe the mechanisms underlying these concepts, and therefore prevents their definition from suffering the same problems faced by the concept of emotions. Motivated behaviors usually fall under one of two categories: approach toward a desired outcome or avoidance of an undesired outcome. Furthermore, motivation can be conceptualized as a two stage process. Once a need arises, and if the desired object is out of reach, the organism engages in an exploratory behavior in search for said object. When the goal object is finally within reach, a consummatory action is performed by the organism to fulfill its desire. It has recently been discovered, that the first stage of any motivated behavior is guided by incentives (LeDoux, 2012). Which means that within the survival circuit theory, any stimulus triggering a circuit also has the potential to drive motivation.

Reinforcement learning was born from and is still used as a standard model to interpret empirical results of classical conditioning experiment. This paradigm has known a rise in popularity within both neuroscience and machine learning. While machine learning researchers adopted variations of this algorithm to robustly solve complex tasks, neuroscientists are trying to map its underlying principles to mechanisms within the brain. So far it has been accepted that dopamine could be interpreted as the reward prediction error signal at the core of the reinforcement learning framework. However, the mechanisms involved in the activation of dopamine neurons are not fully understood yet (but see Sub-Sub-Section 3.3.2 for a possible model). Similarly, more research is needed on the influence dopamine exerts over associative learning, at the basis of classical conditioning. Both problems are explored in more details in Chapter 3. Nonetheless, LeDoux (2012) argues that the survival circuit theory can help shed some light on the systems underlying reinforcement learning in the brain:

> “*The expression of reinforcement as a change in the probability that an instrumental response will be performed may well occur via a generic system in which the reinforcer strengthens the response [. . . ]. But, in addition, survival circuit-specific motivational information is likely to contribute at a fundamental level, providing the stimulus with the motivational value that allows it to ultimately engage the more generic mechanisms that strengthen instrumental responses and that motivate their performance.*” — LeDoux (2012, p. 14)

Activation of any survival circuit leads to adaptive motivated behaviors and learning through association and reinforcement as described so far. However, an additional consequence of the triggering of a survival circuit is the release of hormones in the brain and blood stream. Through their widespread influence the released hormones modulate the neuronal excitability of their target areas. Therefore, they modulate rather than initiate any processing taking place in the receiving brain regions. As described in more details in Chapter 3, dopamine and serotonin are two of the major hormones influencing the different brain systems. In addition to the role they play for reinforcement learning, dopamine raises the sensitivity of neural populations to any incoming stimuli, whereas serotonin inhibits neural activity. As a result, this hormonal modulation implements a simple attention mechanism, whereby stimuli relevant to survival circuits trigger the release of hormones, which in turn sensitize the sensory-motor system further focusing its processing on possibly relevant stimuli. Hence, it creates a sort of positive feedback loop between survival circuits and hormone releasing brain areas. According to LeDoux, the release of hormones into widespread areas of the brain accounts for the global arousal, which accompanies most emotions.

In conclusion, upon entering the brain a relevant stimulus is detected by the survival circuits. This event marks the beginning of what is usually referred to as an ‘*emotional episode*’. The corresponding survival circuit, once activated, simultaneously sends signals to the modulatory, motivational, sensory-motor and other cognitive systems (of relevance in this case are learning and memory components), as well as triggers the autonomic nervous system. The modulatory system releases hormones into widespread areas, placing the brain in a state of generalized arousal, and focusing the attention of the sensory system on similar incoming stimuli. Within the motivational and motor systems, innate and learned adaptive reactions are potentiated, while memories relevant to the survival function are retrieved from both implicit and explicit memory systems. Once the reaction has been performed, new memories are formed and learning occurs. The overall result of survival circuit activity, along with the activation of the memory, learning, motivational and modulatory systems, the bodily feedbacks and the generalized arousal is the establishment of a mental state, which LeDoux calls '*global organismic*' state.

Given the history of the concept of emotions, one might be tempted to equate this global organismic state with an emotional experience, or '*feeling*' as it is usually referred to in humans. However, LeDoux argues that feelings and global organismic states are two independent concepts that can, and in fact do, exist separately from each other. Global organismic states are only the raw material from which feelings, associated with survival circuits activation, are constructed. Such feelings only occur when consciousness: *1*) directly detects the activation of a survival circuit, or becomes aware of the existence of a global organismic state, which itself results from the activity of survival circuits; *2*) appraises and labels this state with an emotion. Other kinds of feelings, that may arise within sufficiently complex brain structures, include those associated with higher-order or social emotions and sensory feelings. Finally, LeDoux (2012) warns against applying introspectively based concepts to other animals. Considering that all humans have the same basic brain, which contains the same basic structures, it is safe to conclude that all humans are conscious creatures and experience emotions in similar fashions. However, since the brain of other species differ from ours, it is impossible for us to know for sure what is the subjective feeling associated with a particular emotion. Or as Nagel (1974) argued, only a bat can experience the world like a bat.

After the failures of the limbic system and the basic emotions theory, neuroscience is nowadays focused on determining the different brain areas that are involved in recognizable emotions, such as fear and disgust, through studies on conditioning, Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI). The circuitry for fear is nowadays very well described (Dalgleish, 2004; LeDoux, 1998) across a variety of species (rats, monkeys and to a lesser degree humans), giving rise to some tentative biologically plausible models of emotions, implemented in networks of artificial neurons. Examples of such models are described in more details in the next two Sub-Sections.

### 2.6.3 Proto-affect for effective functioning

Extending the fundamental ideas behind the OCC model (Ortony et al., 1988), Ortony, Norman, and Revelle (2005) introduced a framework describing a structure around which all information processing in the brain can be organized. They believe that for a system of any complexity, be it an animal, a virtual agent, or a robot, effective functioning depends on the interplay of four domains: *1*) '*affect*', what the organism feels; *2*) '*motivation*', what the organism wants and needs; *3*) '*cognition*', what the organism knows, thinks and believes; *4*) and '*behavior*', what it does. Moreover each of these '*domains of functioning*' need to be considered at three levels of information processing. The first, '*reactive*', level is assumed to be a hard-wired structure, only capable of releasing fixed action patterns. It also plays the role of an interrupt generator for higher levels of processing. The second level, aptly labeled '*routine*' level, is primarily concerned with the execution of automatic behaviors. Its higher mental capacities allow it to form simple expectations about the future, and be the seat of awareness, but not self-awareness. Consciousness and self-awareness are processes that require self-reflective functions, only present at the highest level of processing, also called the '*reflective*' level. Aside from consciousness, the reflective level is the locus of many other complex cognitive processes, such as decision making, assessment of causal relations, complex understanding or abstract reasoning. Although the model is a functional one Ortony et al. (2005) argue that many of its aspects are consistent with neuroanatomical accounts. Therefore, the three levels of information fulfill functions roughly corresponding to: the spine/midbrain/basal ganglia, cortex/cerebellum, and the pre-frontal cortex. This is in a way reminiscent of MacLean's (1970) '*triune brain*' explored above.

The term '*affect*' in this context is used as a superordinate concept encompassing other valenced conditions such as emotions, feelings and moods, but not personality. Emotions, therefore, are simply affect that relates to something, whether this something is an object, an individual, or an event matters not. Further, feelings are taken to be simple readouts of the somatosensory systems, as well as altered states of awareness and attentiveness. According to the authors, any stimulus that the brain encounters will automatically be assigned a value, thus feelings can be considered the most basic form of affect. Full-fledged emotions, in turn, are described as emerging from the cognitive perception of feelings, influenced by motivation and the organism's reactions:

> *"As already indicated, we consider affect to be a general construct that encompasses a wide range of psychological conditions relating to value. However,*

> *even though emotions are more highly specified than other affective states, they do not comprise a discrete category with easily identifiable boundaries. Rather, they vary in their typicality, with some cases being better examples than others. Thus, we propose that the best examples of emotions, which we often refer to as* 'full-fledged emotions'*, are interpretations of lower-level feelings and occur only at the reflective level, influenced by a combination of contributions from the behavioral, motivational and cognitive domains.*" — Ortony et al. (2005, p. 177)

On the contrary to affect that is concerned with value, cognition is defined as '*cold*' processes interested in the stimulus' meaning alone. Motivation, which is usually seen as a part of emotion, in this framework, relates to the different tendencies guiding the organism's actions to accomplish a goal or fulfill its needs. Finally, behavior is simply that, any physical action, whether externally observable or internal (change in heart rate, contraction of the viscera or any other autonomic reaction).

### The reactive level

Ortony et al. (2005) consider the reactive level of information processing as a hard wired structure, only capable of releasing biologically determined reactions to survival-relevant stimuli. Most systems involved at this level can be reduced to simple pattern recognition mechanisms. As a consequence it is rapid and relatively unsophisticated in both its detection capabilities and its behavioral repertoire. Reasons that also make it prone to commit numerous errors, in the form of false alarms (initiating a reaction to a stimulus that is irrelevant) and misses (on the contrary, remaining unresponsive to an opportunity). Finally, due to its simplistic underlying mechanisms and overall lack of mental capabilities, it is assumed that cognition is absent from this level.

A further consequence of the elementary nature of the reactive level, is the fact that the three remaining domains (behavior, motivation and affect) are so closely intertwined that they are better thought of as different characterizations of the same phenomenon. Motivation is reduced to simple drives, which can be viewed as guides to react to changes in the organism's homeostasis. Similarly, motor expressions range from primitive reflexes, to preparatory responses allowing higher levels of processing more time to initiate the next action. Behaviors generated by the reactive level, can be sorted into two broad classes: *1*) approach, resulting from the activation of motor and autonomic systems; *2*) and avoidance, which on the contrary is a consequence of inhibiting the underlying behavioral mechanisms. Finally, the processes related to the domain of affect, are responsible for assigning a value, along the two dimensions of '*positive*' and '*negative*', to every stimulus entering the reactive level. Lacking any of the features attributed to full-fledged emotions, the concept of affect is referred to as '*proto-affect*'. However, its role is not limited to the valuation of stimuli, it also serves the purpose of interrupting ongoing processes and directing the attention of higher levels, toward important events of the environment. Moreover, it is possible for proto-affect to reach higher levels of processing, where it will be interpreted and cognitively extended. Hence, originating full-fledged emotions. Even though, it is deemed simple the reactive level has access to some crude representations of the recent past, enough to allow for habituation and some form of classical conditioning.

### The routine level

Moving up in the hierarchy, leads to the routine level, whose core is the execution of well-learned routines. Having access to elaborate representations of the present and the future, it is capable of a wide range of processing, which extends from conditioning to symbolic processing. According to the authors, it is also the seat of awareness, but not self-awareness. This level initiates and controls both human behaviors and cognition. Since most of its content is unconscious and its processes automatic, they are usually acquired through learning and experience. Even though, the routine level has more mental capabilities than the reactive level, some particular situations still require the control of the reflective-level. For example, in the routine level, cognitive processes can correct for small deviations in expectations. However, if the discrepancy becomes too large, there is a need for the reflective level to evaluate the situation, abandoning the current goal and formulating a new strategy to solve the problem at hand. Through learning and experience the routine level is able to form implicit expectations. Therefore, it is also capable of detecting any violation of those expectations, but not interpret it. This is where the need for a higher level of processing arises.

Cognition finally appears at this level and can be found in the form of automatic unconscious processes, such as perception, categorization, basic language recognition and synthesis to name but a few. Its access to basic awareness combined with a more elaborate concept of affect, allows the organism to have feelings, in the general sense of the term. Although affect has been described as more elaborate, it still lacks a cognitive interpretation and only shares some features with the full-fledged emotions, hence its labeling as '*primitive emotions*'. However, by crossing the two dimensions of positive and negative values, available at the reactive level, with the two levels of time, present and future, represented at this level, four broad categories of emotions emerge:

- a (positive) feeling about a '*good thing*' (present), usually referred to as '*happiness*'.
- a (negative) feeling about a '*bad thing*' (present), which can be translated to '*distress*'.
- a (positive) feeling about a '*potential good thing*' (future), usually labeled '*excitement*'.
- a (negative) feeling about a '*potential bad thing*' (future), often described as '*fear*'.

As a result of the increase in mental capabilities, motivations are not limited to basic drives anymore and can take more varied shapes, such as inclinations, urges or restraints. Furthermore, because information can be stored in memory, motivation can exist even when the cue, that originated it, is not part of the environment anymore. At the routine level, a motivation will persist until it is completely fulfilled. The availability of memory also enables the organism to perform complex actions, since behavioral '*chunks*' can be stored, or learned, and later organized into '*skills*'. On the contrary to the reactive level, at the routine level, the four domains of functioning are related to, but distinct from one another.

### The reflective level

The reflective level is the peak of the information processing hierarchy, suggested by Ortony, Norman and Revelle. It is the seat of advanced cognitive and meta-cognitive skills, as well as consciousness. Any process operating at this level rely to a certain degree on '*reflection*', a special characteristic of higher animals, such as humans or primates. In humans it allows the construction and use of mental models of people, animals, artifacts and of the possible interactions one can have with the other models. More importantly it enables us to infer and represent the thoughts other humans might entertain, about ourselves or other objects (real or abstract). Additionally, humans, in particular, also have access to the notion of self, making it possible to maintain an image of ourselves, have ideals and standards, and introspect. The reflective level is where art is appreciated, abstract reasoning solves mathematical problems and social collaboration, or deception, happen. To put it succinctly, the reflective level is where cognition realizes its full potential.

However, cognition is not the only functioning domain reaching its apogee in the reflective level. Primitive emotions, rising from the routine level, are contextualized, cognitively interpreted and elaborated into full-fledged emotions. Combined with the information present in consciousness, the power of the reflective level enables the rich emotional experiences, supposedly unique to humans. In turn, emotions influence the functioning of both cognition and motivation. According to the authors and the processes supposedly giving birth to full-fledged emotions, the reflective level is where the focus of most appraisal theories lies. Motivation also figures prominently, at the reflective level, and takes the form of long-term goals, plans and strategies that guide behaviors.

On the contrary to all the other domains of functioning, behavior is absent from the reflective level. Based on the fact that the pre-frontal cortex is the anatomical equivalent, in the human brain, to the reflective level, Ortony et al. hypothesize that it has no direct access to somatosensory information. Nor does it have direct control over the motor cortex, or behavior in general. The reflective level is only able to communicate with the lower routine level. As a consequence, cognitive models and representations in the reflective level are based on information rising up from the routine level. All the while behavioral strategies are executed by influencing or controlling the circuitry involved in the routine level.

### Conclusion

The framework introduced by Ortony, Norman, and Revelle (2005) is part of a line of investigation (pioneered by MacLean (1952, 1970)) that is less preoccupied with the concept of emotions and rather focuses on the different interactions between the brain's sub-systems: affect, cognition, motivation and behavior, in this case. It allows scientists, who adopted such a framework, to organize their thoughts, observe the similarities between brain architectures of different animals and easily transfer the knowledge or concepts acquired from one brain structure to another. Other examples of such attempts at organizing the brain into independent yet related systems can be found in the work of Izard (1993), Alexandrov and Sams (2005), and Sloman (1999, 2001).

Although many acknowledge that conscious experience is an important part of emotions, due to the complexity of defining both emotions and consciousness theories are usually confined to the reactive or routine levels, while limited consideration has been given for the role of emotions at the reflective level yet. Consequently, emotions are still restricted to the unconscious part of the mind and interact with the limited cognitive systems that populates it. Certainly, designing a model capable of achieving a level of emotional intelligence on par with the limited form of emotions and cognition found in the routine level would already be quite the accomplishment. However, full-fledged or '*blue ribbon*' emotions as they are referred to by Panksepp (1998) require consciousness and the access to the highest level of cognitive capabilities it affords (Damasio, 1997; LeDoux, 2003; Thagard & Aubie, 2008).

### 2.6.4 Emocon: A model for emotional consciousness

Aware of this issue, Thagard and Aubie (2008) introduced the '*Emocon*' architecture, whose underlying theory seeks to provide a detailed explanation of how conscious emotional experience emerges from the interaction between brain areas coordinated through working memory. According to Thagard and Aubie (2008), a theory of emotional consciousness needs to be able to explain phenomena, such as *1*) emotions '*differentiation*', or the ability to distinguish a wide variety of emotions; *2*) '*integration*', describing the mechanisms by which the outcome of emotional processing is combined, or influences, other cognitive and physiological processes; *3*) the varying degrees of emotional '*intensity*', defined as the degree of mental attentiveness or physical arousal; *4*) the '*valence*', the pleasurable or unpleasurable character of emotions; *5*) and '*change*', the events by which emotional episodes begin and end. Further, the answer should come in the form of hypotheses concerning mechanisms whose outcomes produce said phenomena. Where a mechanism is defined as "*a structure performing a function in virtue of the operations, interactions and organization of its component parts*" (Thagard & Aubie, 2008, p.812). To this end, they identify the crucial components an architecture requires to construct a theory of emotional consciousness as: *1*) representation; *2*) emotional decision-making; *3*) cognitive appraisal; *4*) inference; *5*) and working memory. The main assumption, at the root of Emocon's theory, is that emotional experience is the result of the interactions among all the components, that make up its structure. Furthermore, the concept of emotion is defined as "*a pattern of neural activity in the whole system, including inputs from bodily states and external senses*" (Thagard & Aubie, 2008, p. 817).

Even though, Thagard and Aubie (2008) both had the ambition to implement a neuro-computationally realistic model of emotional consciousness, capable of integrating cognition with physiological and emotional processes, they lacked the computational resources to do so. Instead, they rely on four less complex models: ANDREA, GAGE, EACO and NECO, to provide a '*proof of concept*' of sort. Through the implementation of those models and the experimental results each yielded, Thagard and Aubie are able to describe the mechanisms, and brain regions, underlying the components required for a theory of emotional experience.

#### The GAGE model

The first model part of the Emocon architecture, has been published by Wagar and Thagard (2004). It is named GAGE in honor of the late Phineas Gage (see Section 2.6 and (Damasio, 2008) for a detailed account of Phineas Gage's case), whose disability it intends to model.

According to Wagar, the overall goal is to present a neuro-computational account of the mechanism underlying emotional reaction. On the contrary to Armony's model (see Section 4.1) the author intends to describe the system through a neuroanatomically realistic model. Therefore, the basic processing units are represented by artificial spiking neurons. Using these neurons as basic component of the network, also serves to highlight an important factor in the production of emotional reactions: time. Wagar and Thagard, adopted the '*somatic marker*' hypothesis depicted by Damasio (2008) as foundation. GAGE expands on it to describe the gating mechanism implemented by the nucleus accumbens (NAcc). The NAcc in this perspective is seen as integrating cognitive information from the ventro-medial pre-frontal cortex (VMPFC) and hippocampus, with emotional data produced by the amygdala.

The somatic marker hypothesis (Damasio, 2008; Wagar & Thagard, 2004) posits that, the system made up by the VMPFC and amygdala interact through a set of inter-connections to produce said somatic markers. The VMPFC provides feasible behavioral options for the current situation. These options are then processed in the amygdala through '*As if*' mechanisms that evaluate the expected outcome for each of them. This evaluation in turn is sent back to the VMPFC, where it is stored. After that, if a set of stimuli elicit an already marked behavioral option, the VMPFC sends both the option and its corresponding, retrieved, somatic marker to higher level decision making processes and/or motor effectors. The somatic marker system, as introduced by Damasio, allows the brain to quickly sort through the set of feasible reactions for the current situation. Assuming that the brain always selects the most adaptive and beneficial option, the somatic markers serve as an indicator of which options can be safely ignored (those marked as having an unpleasant outcome), and which ones require more consideration (those marked as having a pleasant outcome):

> "*In this perspective, somatic markers make the decision process more efficient by narrowing down the number of feasible behavioral alternatives, while allowing the organism to reason according to the long-term predicted outcomes of its actions.*" — Wagar and Thagard (2004, p. 68)

GAGE builds on this mechanism and seeks to show that the integration performed by the NAcc, works as a gating mechanism, which lets through only those options that are consistent with the current context. As a consequence, neurons are organized into five anatomically recognized groups, corresponding to the brain areas of the VMPFC, amygdala, hippocampus, NAcc and ventral tegmental area (VTA). The mechanism the NAcc is theorized to use for gating information is described by Wagar and Thagard as follow:

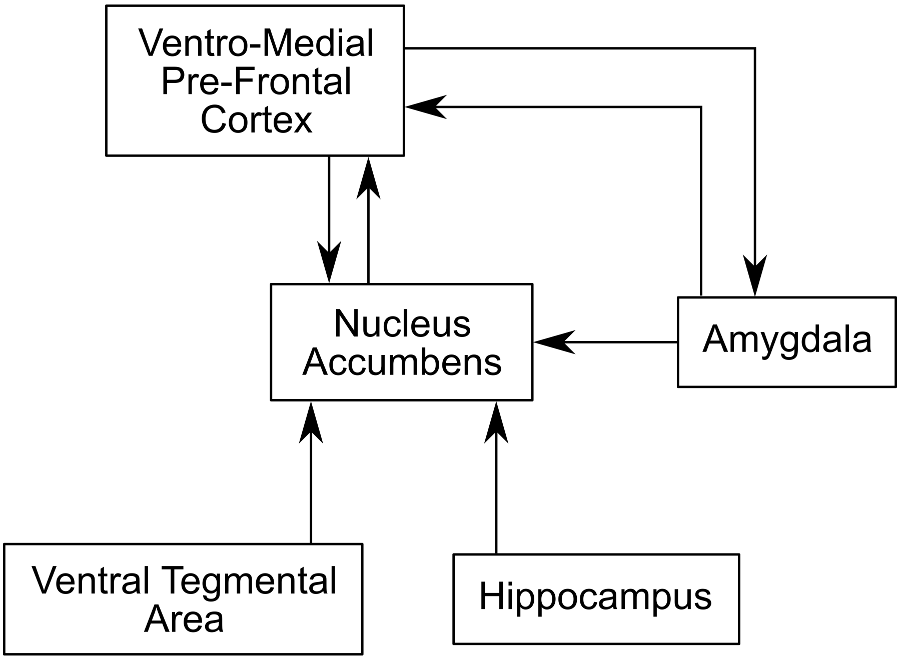


Figure 2.10 *This figure shows the different brain areas included in the GAGE architecture. Introduced by Wagar, this model focuses on the role of the nucleus accumbens, which is hypothesized to implement a gating mechanism. Based on the* 'somatic marker' *hypothesis suggested by Damasio, the ventro-medial pre-frontal cortex (VMPFC) sends plausible options to the amygdala, which evaluates their expected outcomes, via* 'as if' *mechanisms. The value of each option is then sent back to the VMPFC and stored as a somatic marker. In case such an option presents itself again, in the future, its somatic marker will be retrieved and sent to the nucleus accumbens alongside the option. According to Wagar, for an option to be allowed through, it needs to be assessed as having a desirable outcome by the amygdala, as well as make sense in the context provided by the hippocampus. Only in this situation, is it possible for the activity within the nucleus accumbens to overcome the heavy inhibition maintained by the ventral tegmental area.*

Source: Wagar and Thagard (2004, p.69), used with permission from the American Psychological Association, Inc.

- The VMPFC, in line with the somatic marker hypothesis, provides potential responses to and appraise the current situation.
- The amygdala, processes the somatosensory information associated with the current situation, and produces valenced bodily states. Through interaction with the VMPFC, the amygdala also evaluates the expected outcomes of the behavioral options.
- The hippocampus brings contextual information to the NAcc.

Further, whereas VMPFC and amygdala neurons are characterized by low-amplitude short spike trains, hippocampal units fire high-amplitude long-lasting spikes. However, on its own, neither the VMPFC, amygdala nor the hippocampus, is capable of causing the NAcc neurons to fire, since they constantly receive heavy inhibitory projections from the VTA. Leaving the NAcc in a perpetual hyperpolarized state. Therefore, for a behavioral option to be able to go through the NAcc, it is required that all three areas (VMPFC, amygdala and hippocampus) fire synchronously, and that the option is coherent with the context. This means that the pattern of activation sent by the VMPFC, representing an option, has to overlap with a sub-population of the NAcc neurons depolarized by hippocampal output.

To confirm the integration capabilities of the NAcc, GAGE has been implemented and used in two simulations. The computational implementation of the GAGE model used a total of 700 spiking artificial neurons. Each individual unit was modeled after a single-compartment leaky integrate-and-fire neuron.

The first experiment implemented by Wagar and Thagard, simulated the '*Iowa Gambling task*'. In its original form, the Iowa gambling task, has been developed by Damasio (2008) to detect damage to the VMPFC, the area through which Phineas Gage's tamping iron is theorized to have pierced. For the task proper, a subject is presented with four decks of cards. He is then asked to draw a card from any of the decks, until he is told to stop. For each deck there is an independent probability that drawing a card will result in a reward, rather than a punishment. Therefore, by manipulating the probability for reward, it is possible to have '*good*' decks and '*bad*' decks. The subject can switch decks as many time as he wants, and whenever it pleases him. At the beginning of the experiment, the only unknown element for the subject is the quality of each deck (as good or bad). While developing his somatic marker hypothesis, Damasio established that individuals with intact pre-frontal cortex, were able to learn, through experience, to draw cards from only the good decks, and avoid the bad ones. On the contrary, individuals with damaged VMPFC, were unable to take advantage of the long-term expected outcome provided by the interaction of the VMPFC and amygdala. Therefore, they focused more readily on the short-term rewards, and ended up constantly choosing the bad decks. For the computational simulation, used to test the GAGE model, only two decks of cards were used (one good, the other bad). The inputs presented to the VTA and hippocampus were fixed, meaning that the current context was maintained throughout the simulation. Each turn, the VMPFC had the choice between drawing a card from either decks. However, the good deck was associated with a pleasant bodily state presented to the amygdala, whereas the bad deck was tied to an unpleasant state. The results gathered over several iterations of the Iowa gambling task, using GAGE, showed that the model was able to reproduce the emotional reactions observed for both intact and damaged human subjects. Hence, proving that the NAcc is capable of integrating cognitive information from the VMPFC with emotional data from the amygdala.

The second experiment simulated the study performed by Schachter and Singer (1962) (see 2.3 for a brief description). While for the first experiment the context was maintained by presenting the hippocampus with a fixed pattern of inputs, in this second simulation, the amygdala had a static input. The hippocampus was presented with two types of input, one corresponding to a pleasant context, the other to an unpleasant one. In this case the VMPFC did not represent behavioral options for a given situation, but assumed its second role, that of physiological arousal. Similar to the hippocampus, the VMPFC was presented with either a pattern corresponding to anger or one symbolizing euphoria. Again, after several iterations, the data gathered showed that GAGE, was able to label an emotion as either anger or euphoria, based on the context. Thus, proving that the NAcc is also capable of integrating physiological appraisal, with contextual data.

In Wagar and Thagard's perspective, although GAGE has proved capable of integrating cognitive information with emotional reactions, as well as account for the influence the current context exerts on this process, the model remains limited in its predictive power. As it stands, GAGE is able to faithfully reproduce results observed on the Iowa gambling task, for both patients with and without damaged VMPFC. However, according to the current model, a lesion to the NAcc would see the whole mechanism crumble. In reality, due to its plasticity, the brain would instead be able to recover and relocate part of its abilities. Furthermore, because of the restricted number of brain areas implemented, GAGE accurately models the mechanism underlying the production of covert emotional reactions. Yet, it is still unable to explain the processes through which high-level decisions, or cognitive appraisal, happens. In spite of those limitations, Wagar and Thagard still consider GAGE to be a successful first step in using computational neuroscience to help further understanding of human decision making and the impairments that accompany injuries to the pre-frontal cortex.

### The ANDREA model

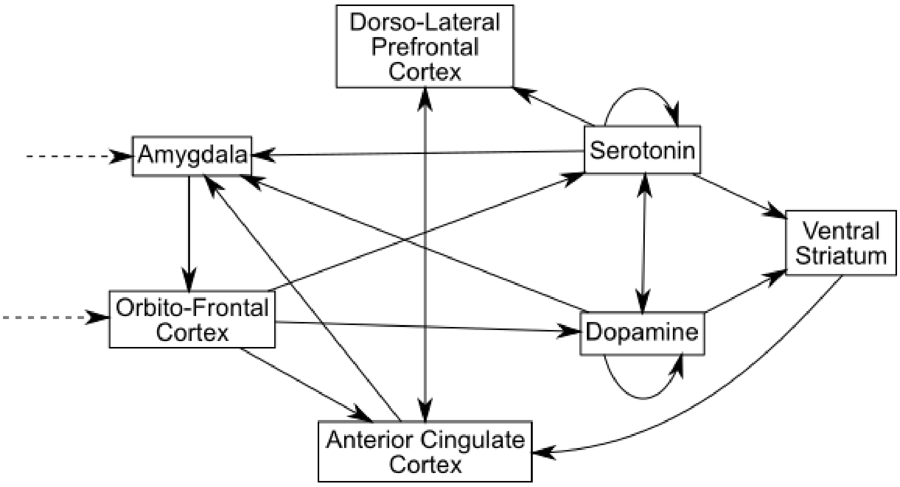


Figure 2.11 *The ANDREA model is an attempt at describing the neural mechanisms underlying motivation, cognition and affect. Within this architecture, both the amygdala and the orbito-frontal cortex (OFC) receive external stimuli. The amygdala uses this information, as well as signals coming from the dopamine (DA) and serotonin (5-HT) systems, to derive the current state of affective arousal. According to Litt et al; the OFC is responsible for perceiving the punishing or rewarding nature of external stimuli. Its output is modulated by the amygdala's activity. The OFC then projects to the DA and 5-HT systems, as well as to the anterior cingulate cortex (ACC). On the contrary to TD-learning, which focuses solely on the DA system, ANDREA employs an opponent mechanism to determine the error in the prediction of rewards and punishments. Of the last two areas, the dorso-lateral pre-frontal cortex is in charge of representing and selecting goal-directed behavior, while the ACC, given its central position, is tasked with detecting any conflict between the current behavior and the active goal.*
Source: Thagard and Aubie (2008, p.815), reproduced with permission from Elsevier.

ANDREA is another neuroanatomically realistic model, introduced by Litt, Eliasmith, and Thagard (2006, 2007, 2008). It was born as a result of Litt, Eliasmith and Thagard noticing that, while imaging studies and neurophysiological experiments had greatly advanced our understanding of the brain structures underlying the interactions between motivation, affect and cognition, the description of the mechanisms by which these structures operates had lagged behind. Further, apart from the reinforcement learning paradigm (detailed later in Chapter 3), most existing models of the reward processing system, were limited to the task-specific level of explanation. Litt et al. believe that a larger-scale synthesis is required to move beyond and produce a truly global model of motivated behavior. A growing wealth of evidences (McClure, York, & Montague, 2004; Schultz, 2000), points toward the involvement of brain areas such as the amygdala, orbito-frontal cortex (OFC), ventral striatum (VS), anterior cingulate cortex (ACC) and dorso-lateral pre-frontal cortex (DLPFC) in reward processing and related goal representation, for humans and higher-order primates. Therefore, by implementing a neuroanatomically realistic model of these structures, as well as their diverse interactions, ANDREA constitutes a first attempt at providing a broad mechanistic description of the human reward processing system. Moreover, as a consequence of its integration of the current theories about how the different reward processing tasks are performed, ANDREA also provides a novel and detailed neural basis for '*loss aversion*': the fact that animals usually prefer avoiding losses, rather than expanding a lesser amount of energy to achieve equal gains.

As a computational model, ANDREA has been implemented with a total of 7600 spiking leaky integrate-and-fire artificial neurons, split into seven neural populations, as shown in Figure 2.11. The amygdala, OFC, VS, ACC and DLPFC were represented by artificial networks with varying numbers of neurons (in the range $[800, 1200]$), while the dopamine (DA) and serotonin (5-HT) systems were each modeled by 1200 neurons spread across several discrete sub-populations. ANDREA relies on the '*Neural Engineeting Framework*' (Eliasmith & Anderson, 2003) for the actual implementation of its basic processing units.

Recent theories of emotions, commonly attribute to the OFC, the role of perceiving the punishing or rewarding nature of external stimuli. Further, according to the authors, its processing is influenced by the current context (not represented in ANDREA), and more importantly, by its interactions with structures involved in emotional processing (Damasio, 2008; Litt et al., 2006, 2008). One such brain area is the amygdala, which for a long time has been considered as the processing center for negative emotions, such as fear or anger. However, this perspective has recently been challenged by McClure, York, and Montague (2004), who argues that amygdala output activity is rather related to a stimulus' salience. Additionally, findings from Adolphs et al. (2005) indicate that the amygdala might also play a role in '*attention*'. Therefore, within the ANDREA model, the amygdala is used to represent the individual's state of affective arousal. Additionally, its interaction with the OFC, to modulate the valuation process of a stimulus, is based on evidence from vision research, which points in the direction of the amygdala having a multiplicative scaling effect of attention on the observed saliency of visual inputs (Treue, 2001). As a consequence, if $V$ denotes the orbito-frontal valuation of a given stimulus, and $A$ the emotional arousal, encoded by the amygdala, then the modulated valuation, at time $t$, is given by the equation:

$$V^*(t) = V(t) \times A(t) \tag{2.1}$$

Thus, for low levels of amygdala activity, the outcome of OFC processing will be attenuated. On the contrary, high levels of activation, results in an amplification of stimuli valuations.

The dopaminergic neurons, of the ventral tegmental area and subtantia nigra pars compacta, have for a long time been implicated in the processing of '*reward prediction error*'. Activity of these brain regions have been the target of the reinforcement learning paradigm (see Chapter 3 for more details), from Rescorla and Wagner (R. Rescorla, 2008; R. A. Rescorla & Wagner, 1972), to its resurgence in the hands of Sutton and Barto (Sutton, 1984; Sutton & Barto, 1998). The dominant approach developed within this framework, the '*temporal difference*' (TD) learning, computes the reward prediction error ($E$) based on the difference between the latest reward valuation and a weighted sum of all previous rewards ($P$). By substituting the current valuation, with the modulated signal of the OFC, expressed in Equation 2.1, the predicted error in reward could be expressed as:

$$E(t) = V^*(t) - P(t-1) \tag{2.2}$$

$$\text{With} \quad P(t) = P(t-1) - \alpha \times E(t) \tag{2.3}$$

Where $\alpha$ is a constant representing the learning rate. As a result, for each positive reward prediction error, neurons in the DA population will fire for a short amount of time with an amplitude proportional to the discrepancy between the predicted and the actual reward. On the contrary, for each negative reward prediction error, activity will dip belove baseline level. Even though, TD-learning, and by extension the reinforcement learning paradigm, is the most widely accepted model to account for reward prediction error, it has been challenged by Daw, Kakade, and Dayan (2002). They claim that due to the low baseline firing rate of the dopamine system, a DA-only scheme is unsuitable for computing highly negative prediction errors. Furthermore, Daw and colleagues, suggest the serotonin system to be interacting in opponency with the dopamine system. It is this new perspective that Litt et al. chose to implement within ANDREA. As a consequence, the DA neural population reacts to appetitive stimuli, whereas the serotonin system is primarily triggered by aversive stimuli. Finally, encodings from both the dopamine and serotonin systems are consolidated in the VS, which is also part of the reward prediction error mechanism. Within ANDREA, the main target of the opponent system is hypothesized to be the amygdala. With both the dopamine and serotonin systems feeding back to the amygdala, its output activity can thus be described by:

$$A(t) = A_e(t) + \beta \times DA(t) + \gamma \times 5 - HT(t) \tag{2.4}$$

Where $A_e$ is the base arousal level determined by external stimuli unrelated to reward. It is important to note that this expression allows for an asymmetry in the influences exerted by the dopamine and serotonin systems ($\beta \neq \gamma$).

The ANDREA model is completed by two of the most important brain areas: the DLPFC and ACC. The DLPFC has been deemed crucial, by Litt and colleagues, for the planning, representation and selection of goal-directed behavior. In the human brain it is theorized to make use of the outcome of ACC processing, as a basis for its own reasoning mechanism. However, to limit the scope of their theory and keep the model's complexity to a manageable level, the authors chose to leave this part aside. Instead, they propose a raphe-dorsolateral-cingulate-amygdala pathway by which negative prediction errors can further influence arousal. Due to its central position, receiving projections from the OFC, DLPFC and VS, the ACC plays an important role in emotional consolidation and detecting conflicts between current behavior and desired results. Adopting this perspective for its implementation inside ANDREA, Litt et al. further suggest feedback connections to the amygdala that help explain psychological and neuro-imaging results. The ACC is tasked with analyzing the results from orbito-frontal processing (modulated stimulus valuation), to produce an appropriate behavioral response: approach, if the stimulus has a positive value; withdrawal, otherwise. The behavioral valence is further integrated with the reward prediction error, from the VS, before being fed back to the amygdala and projected to the DLPFC. As a result, if a given behavior is concurrent with a positive prediction error, the individual is rewarded for his action and that behavior is strengthened. On the contrary, in the case of an action occurring simultaneously with a negative prediction error, conflict and the potential for behavior modification arise. The individual is punished and a weakening of the behavior is modeled via attenuation of the cingulate cortex's activity. Labeling $C$ the cost for conflict and potential behavior modification, the global activity of the amygdala population is described by:

$$A(t) = A_e(t) + \beta \times DA(t) + \gamma \times 5 - HT(t) + C(t) \tag{2.5}$$

Consequently, in ANDREA, the loss aversion phenomenon, is accounted for by the calibration of the asymmetric contributions of the dopamine and serotonin systems to amygdala output ($\gamma > \beta$ in Equation 2.5). In turn, amygdalar activity modulates OFC processing outcome, through the multiplicative scaling realized in Equation 2.1. As a consequence, larger values of $\gamma$ result in a bias towards negative stimulus evaluation. Litt et al. justify this particular calibration by saying that:

> "*Such disparity might have an evolutionary basis: negatively appraised events may often jeopardize survival and further reproduction, while the same vital saliency does not commonly accompany boons.*" — Litt et al. (2006, p. 497)

Additionally, losses may further bias arousal via the cost for conflict and behavior modification (denoted $C$ in Equation 2.5), whose value results from the activity of the pathway originated by the dorsal raphe nucleus (seat of the serotonin system). The amygdala playing an important role in attention, in addition to processing state of affective arousal, means that any conflicting stimulus, which may require a shift in the current strategy, will easily draw attention upon itself. Therefore, through the different feedback loops to the amygdala and its interaction with OFC, aversive external stimuli will be overvaluated. Litt et al. conclude their analysis of the loss aversion mechanism modeled in ANDREA as follow:

> "*The nature of loss aversion is thus well replicated by our [the ANDREA] model, which proposes the first detailed, biologically realistic neural basis for the phenomenon.*" — Litt et al. (2006, p. 498)

Finally, a lesion study has been simulated to investigate the consequences of ablating the serotonin system, for the activity of the different mechanisms described so far. When the raphe nucleus is lesioned, negatively appraised stimuli can no longer increase arousal. Since the brain region originating the raphe-dorsolateral-cingulate-amygdala pathway is cut off from the system, the cost for conflict $C$ disappears from the output activity of the amygdala, leading to the disappearance of the loss aversion. Further, the direct feedback from the serotonin system, to the amygdala also wanes. As a result, stimuli negatively evaluated by the OFC, do not benefit from either attenuation or amplification. The overall outcome for the whole model is that the individual indulges in more risky behaviors, disregarding the dire consequences some actions might have. Litt and colleagues, conclude that:

> "*The outcomes of this lesioning agree well with the findings of experimental studies, …*" — Litt et al. (2006, p. 498)

Moreover, they urge other scientists to use the ANDREA model, or models of their own, to investigate the effects of ablating the dopamine system, as a complementary study to the one presented here. Such experiments would also provide with a concrete understanding of the behavioral consequences observed in animals.

In conclusion, by synthesizing wide ranging psychological experimental data, as well as explaining the computation happening both between and within each brain region, the ANDREA model proved to be useful in providing researchers with a detailed explanation of the neural structure underlying behavior motivated by affective stimulus valuation. Yet, there remains much to explore within smaller subsets of the reward system, the complex mechanisms upon which the DLPFC relies to make concrete decision, and the different influences the former system has on the latter.

### The EACO model

Although dimensional theories usually differentiate emotions based on their valence and arousal properties (Mehrabian, 1980, 1996; Scherer et al., 2010), Thagard and Aubie (2008), in a fashion similar to J. a. Russell (2003), argue that those dimensions are not sufficient for an individual to be able to make out the difference between, for example, sadness and anger (both negatively valenced and with intensities ranging from moderate to extreme). Therefore, for fine discrimination of the full range of emotions, cognitive appraisal is needed.

Since the definition of emotion, used within the context of the Emocon theory, is similar to the one Scherer suggested in the Component Process Model (CPM, see Subsection 2.4.1 and (Sander et al., 2005; Scherer, 2009; Scherer et al., 2010; Scherer & Ekman, 1984; Scherer et al., 2001) for details), and because the conceptual mechanisms, by which an individual transition from one emotional state to the next, are also comparable, the CPM was chosen as an inspiration for cognitive appraisal. However, rather than adopting the established sequence between the Stimulus Evaluation Checks (SECs), Thagard and Aubie (2008) claim that a similar or better appraisal mechanism can be implemented using a parallel constraint satisfaction model. As a consequence, the Emotional Appraisal as COherence (EACO) structure combines the parallel satisfaction model of coherence introduced by Thagard (1989) with the theory underlying the CPM to produce a computational model of emotional appraisal in the brain.

EACO has been implemented, as an artificial neural network, containing 67 units, spread across three layers. For each time step, the activation value of unit $j$ is updated with the following equation:

$$a_j(t+1) = a_j(t) \times (1-d) + enet_j \times (max_j - a_j(t)) + inet_j \times (a_j(t) - min_j) \tag{2.6}$$

Where $d = 0.05$ is the decay factor, $max_j$ and $min_j$ are the maximum and minimum activation level for unit $j$, respectively, and the net excitatory and inhibitory inputs are given by:

$$enet_j = \sum_i \omega_{ij} \times a_i(t) \quad \text{for } \omega_{ij} > 0 \tag{2.7}$$

$$inet_j = \sum_i \omega_{ij} \times a_i(t) \quad \text{for } \omega_{ij} < 0 \tag{2.8}$$

With $\omega_{ij}$ the strength of the connection between sending unit $i$ and the receiving unit $j$. The first layer, is made of a single '*special*' node, which maintains an activation level of 1 at all time. The second layer, comprised of 52 nodes, represents the 16 SECs. The value of each individual SEC is encoded across a different number of nodes depending on the values it can assume. For example, the '*low familiarity*' of a situation is represented by a node different from the one encoding for a '*high familiarity*'. The third, and final layer, contains 14 nodes each standing in for an emotion chosen among: happiness, joy, disgust, contempt, sadness, despair, worry, fear, irritation, rage, boredom, shame, guilt and pride. Units are connected, between layers through bidirectional excitatory projections, and to each other within the same layer via inhibitory links. As a result, the strongest emotion, in the last layer, gains full activation, while inhibiting all other emotions. However, for two emotions with similar SECs, for example happiness and pride, even if happiness is the most activated unit and inhibits the other emotions due to the similarity in SECs pride can still overcome the lateral inhibition and activate a small amount. This way, EACO also accounts for instances of mixed emotions.

Even though EACO has been implemented in a neural network, to take advantage of the natural parallel processing capabilities of such structures, as well as allow for partial, or mixed, emotion recognition, it is still limited by its use of localist representation, where each emotion is modeled by a single unit, a far cry from the encoding capabilities a fully distributed system could achieve. Consequently, Thagard and Aubie next suggested the NECO model.

### The NECO model

Thagard and Aubie employed the Neural Engineering Framework (NEF), of Eliasmith and Anderson (2003), to develop a more realistic model of parallel constraint satisfaction, which has been shown to be able to reproduce results realized by the explanatory coherence (Thagard, 1989), an equivalent localist model. Furthermore, they sought to explain how biologically realistic neural populations, organized in recognized brain regions, including the lateral pre-frontal cortex (LPFC) and amygdala, can compute a kind of emotional coherence through interactions.

As implemented by Thagard in 1989, explanatory coherence solves a parallel constraint satisfaction problem, such as theory evaluation, by representing each proposition with a single unit, whose activation level is a real number constrained between $-1$ and 1. Nodes are connected via positively or negatively wighted links, indicating explanatory or contradictory relations, respectively. Building on the example of theory evaluation, we have two pieces of evidence $E_1$ and $E_2$, as well as two incompatible hypotheses $H_1$ and $H_2$. With $H_1$ able to explain both $E_1$ and $E_2$, while $H_2$ can only account for $E_1$. The resulting network is depicted in Figure 2.12. With the nodes representing both evidences connected to a '*special*' unit, whose activation level is maintained at 1. To know which of the hypotheses is stronger for our theory, we simply let the network update the activation values of the node representing $H_1$ and $H_2$, at each time step. Once the values of the hypotheses have stabilized, it is expected in this case for the node corresponding to $H_1$ to have a high positive value (close to 1, the maximum allowed), whereas the second node, tied to $H_2$ should have a rather low value, close to 0 or less depending on the strength of its connection with the first evidence unit. In 1989, Thagard implemented ECHO, a localist model, which has been shown to be able to corroborate the coherence of the '*Theory of Evolution*', using hypotheses and evidences similar to the ones Darwin had at his disposition, in his time. Using the NEF, Thagard and Aubie extended the ECHO model, into a parallel distributed and biologically realistic model, which they called NECO.

Whereas ECHO used a single unit to represent a proposition, NECO uses 1000 spiking neurons to encode the acceptability levels of all propositions. In its first iteration, NECO employed overlapping neural populations to represent the acceptabilities of multiple propositions. Hence, combined the populations stand for a real number vector space of $n$ dimensions, where $n$ propositions are being considered. As a consequence, each dimension is representative of the acceptability of one proposition. To be able to store multiple real numbers in a single population, NECO randomly tunes each neuron, such that its spiking rate correlates with a direction, at the beginning of every experiment. Consequently, some neurons can be tuned in between multiple dimensions, and therefore are excited by multiple propositions. The fact that a single neuron can contributes to the acceptability of two different proposition, is what makes NECO's first iteration, a truly distributed system. Coherence and incoherence between propositions, in NECO, are represented by recurrent excitatory and inhibitory connections, respectively, between the neurons encoding for the corresponding dimensions. Thus, if one of two coherent propositions is acceptable, its neural representation will also excite, to some extent (controlled by synaptic weights), the neural population encoding for the acceptability of the second proposition. The reverse being true for contradictory propositions. Similar to localist models, such as ECHO, parallel constraint satisfaction is achieved by externally stimulating the accepted evidences, and letting the firing patterns of all populations settle into a stable state. The resulting patterns, once decoded with the appropriate function, represent the acceptance or rejection of each proposition, and have shown to maximize the constraint satisfaction (Thagard, 1989). NECO's first implementation has been shown to be able to reproduce the behavior of ECHO, through a series of simulated theory evaluations, albeit in a more neurologically realistic manner.

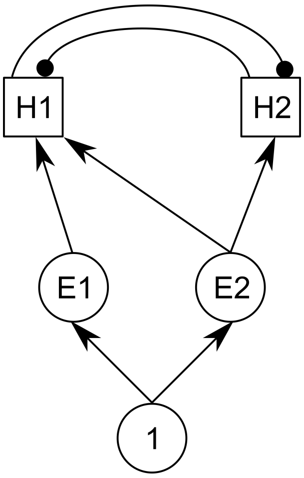


Figure 2.12 *Here is an example of a localist model of explanatory coherence, introduced by Thagard. The goal of this kind of model is to discover the best hypothesis, given a set of evidences for each of them. In this case, there are two hypothesis, labeled* $H1$ *and* $H2$ *at the top. As indicated by the inhibitory connections in-between those hypothesis, they are mutually exclusive. The first evidence (*$E1$*) can only be explained by the first hypothesis, whereas the second one has relevance for both hypothesis. A fact that is clearly indicated by the excitatory connections between the respective nodes. After the activity of all nodes within the network have stabilized,* $H1$ *will be selected as the best hypothesis. More complex networks, can not only select the best hypothesis, but also reject any evidence that do not fit.*
Source: Thagard and Aubie (2008, p.823), reproduced with permission from Elsevier.

Through another model, called HOTCO, Thagard (2002) was able to show that if each unit representing a proposition had a numerical valence, in addition to its standard acceptability value, then the interactions between cognition and emotion could be understood in terms of parallel constraint satisfactions. Due to Thagard and Aubie's willingness to make NECO neurologically plausible, valence was modeled through interactions between multiple brain areas, rather than as a function of a single neural population, which also had to encode for acceptability. Following the somatic marker hypothesis introduced by Damasio (2008), as well as some of the ideas suggested by the GAGE model (Wagar & Thagard, 2004), in NECO it is assumed that each proposition carries an emotional memory. This memory is later used to elicit emotional responses, when and if the proposition reaches working memory.

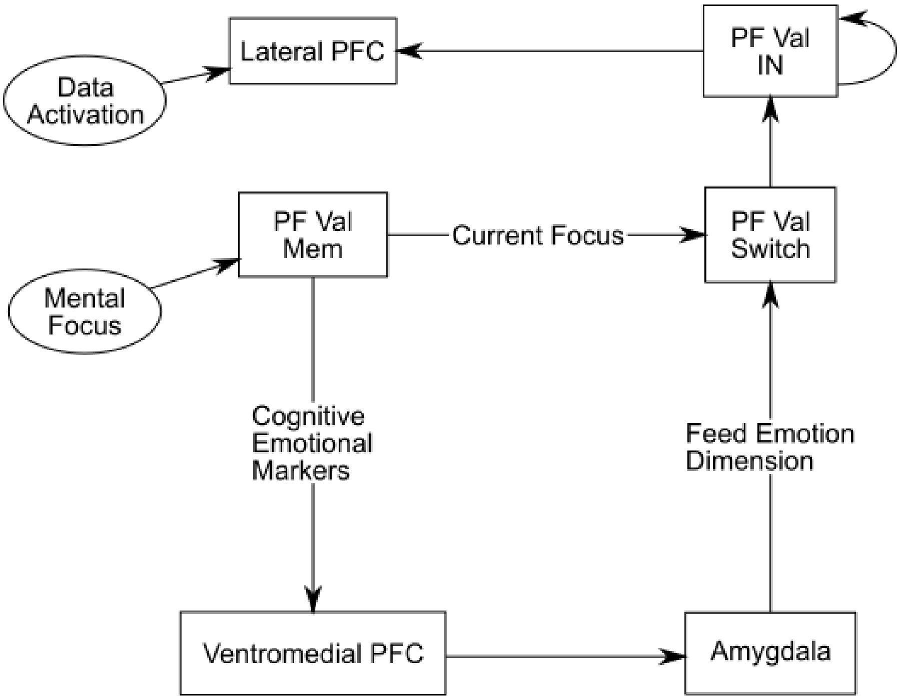


Figure 2.13 *NECO is a neurologically realistic model illustrating how emotions could be construed as the solution to a parallel constraint satisfaction problem. The lateral pre-frontal cortex (lateral PFC) computes the acceptability levels of all propositions, within a single neural population. The PF ValṀemàrea is a long-term storage for the emotional value of each proposition. It sends the emotional markers of all propositions to the ventromedial PFC, where they are consolidated and summed together. The output is projected to the amygdala, which transforms it into an emotional signal. PF ValṀemàlso connects to the PF ValṠwitch area, sending it the current mental focus. In turn, PF ValṠwitch encodes both the emotional signal and mental focus and forwards their product to PF Valİn, where the information will be maintained via recurrent connections. Through its link to the lateral PFC, the PF Valİn area* 'pulls*' the acceptability levels of each propositions toward their respective emotional markers.* Source: Thagard and Aubie (2008, p.827), reproduced with permission from Elsevier.

The second iteration of the NECO model, developed to fill some of the gaps present in the Emocon theory, organizes its neural populations into six interacting areas shown in Figure 2.13: the lateral pre-frontal cortex (LPFC), the ventro-medial pre-frontal cortex (VMPFC), the amygdala, and three synthetic pre-frontal cortical areas labeled '*PF Val. Mem.*', '*PF Val. Switch*' and '*PF Val. IN*', where '*Val.*' stands for valence. Additionally, two external input stimuli are provided to the model, namely the '*Data Activation*', which projects directly to the LPFC and represent the outcome of all cognitive processes. The '*Mental Focus*', second external stimulus, in this instance plays the role of working memory, and triggers the activation of the emotional memory for a given proposition, stored in PF Val. Mem. To save on computational power and complexity, rather than modeling each of the six regions with a combined neural population for all propositions, this instance of NECO uses distinct sub-populations to encode the properties of each individual proposition within an area. Consequently, the LPFC contains the acceptability values of each propositions. The emotional memory, or somatic marker (Damasio, 2008), characterizing a proposition is stored in the PF Val. Mem. area, which in turn projects to the VMPFC. As described in the GAGE model, the role of the VMPFC is to consider all plausible behavioral responses, sorting or rejecting them based on their expected valenced outcome (Wagar & Thagard, 2004). In NECO, the VMPFC consolidates the emotional valence input for all propositions and sends the total sum of these inputs to the amygdala. Next, the amygdala integrates the emotional valences into a global emotional state, which it forwards to the PF Val. Switch. In turn the PF Val. Switch, outputs the '*product*' of the focus value, for a given proposition, with the amygdalar emotional state. So that only the currently active proposition is associated with the emotional state. Once the PF Val. IN population receives a value, this information is maintained via recurrent connections, keeping it active in working memory as well. Through its direct connections with the LPFC, the PF Val. IN component virtually '*pulls*' and '*pushes*' the acceptability of the currently active proposition, by an amount proportional to its emotional marker. As a result, positively marked propositions are strengthened and more likely to be chosen, whereas negatively marked ones gradually fade away.

Based on the results gathered through a series of simulations, Thagard and Aubie conclude that, far from implementing the complete Emocon theory, the NECO model remains of importance as it provides plausible explanation for how emotional appraisal construed as parallel constraint satisfaction can plausibly be implemented using distributed representations and inter-connections with emotionally important brain regions, such as the VMPFC and amygdala. Therefore, it partially fills some of the gaps that remain in this complete account of emotional consciousness.

### Conclusion

In conclusion, by combining the many neural components and mechanisms, described so far, we get an integrated system capable of accounting for the wide range of components and phenomena crucial to the emergence of emotional consciousness:

- Representation and working memory: For the brain to be able to do more than simply react to sensed aspects of the external environment or the internal state of the body, encoding needs to be realized at different levels of complexity. At the lowest level, if an external or internal stimulus manages to activate a sensory organ a signal is sent to the brain or more specifically to a neural population. In turn, neurons within this population excited by the incoming signal fire following a causally correlated pattern. This pattern which is incrementally built by tuning the neural population via statistical dependency to the neural signal coming from the senses, represents a feature of either the world or the bodily state detected by one of our many sensory devices. For higher-levels, Eliasmith and Anderson (2003), have been able to describe how neural populations can not only build representations of the firing patterns of other neural populations directly projecting to them, but also transform these encodings in complex ways. Hence, introducing the possibility to form representations of representations, and allowing the brain to develop higher-order thoughts and processes. One such high-order process is working memory, which involves the integration and short-term storage of differently encoded information, as well as the executive processes of attention and task selection. While the representation component is ubiquitous in an architecture made of interconnected populations of neurons, working memory mainly relies on the dorso-lateral pre-frontal and anterior cingulate cortices, as has been described in the ANDREA model.
- Emotional decision making: The different mechanisms by which emotion is able to influence the decision-making process have been successfully modeled by both the GAGE and ANDREA architectures. However, at least two other brain regions, the thalamus and insula, are thought to be of importance for emotional consciousness, due to their extensive connectomes, especially their links to both internal and external sensors. The thalamus, considered the gateway to the brain, receives projections from all external sensors, save for the olfactory bulb which is directly connected to the amygdala, and sends efferent connections to the amygdala and the cortex. The insula integrate somatic information, which it passes along to other cortical areas.

- Inference and appraisal: As it stands, the Emocon model is still missing a detailed account of the mechanisms underlying appraisal and inference. However, the EACO and NECO models both represent attempts at solving this problem, albeit at different levels of complexity and details.

Moving on to the emotional phenomena:

- Valence: The concept of valence in this particular architecture, can be considered at two levels. At the level of the stimulus, as explained by the ANDREA model, the orbito-frontal cortex plays a major role in perceiving the pleasurable/unpleasurable nature of any given stimulus. While, when applied to a mental state, valence is then accounted for by the opponent system, comprised of the dopamine and serotonin neural populations. Further, imaging studies have found that, in addition to the dopamine and serotonin systems modeled by ANDREA, positively valenced states are correlated with the activation of the left pre-frontal cortex, whereas negative mental states are associated with increased activity of the right pre-frontal cortex.
- Intensity: The most natural explanation for a difference in intensity, for two similarly valenced emotions, is in terms of firing rates in the relevant neural populations: dopamine system and left pre-frontal cortex, for positive states; serotonin system and right pre-frontal cortex, in the case of negatively valenced mental state. Furthermore, Anderson et al. found that amygdala activation is related to emotional intensity, but not valence. On the contrary, the orbito-frontal cortex react to the valence of an emotion, regardless of its intensity (A. K. Anderson et al., 2003). Thus, corroborating the mechanism hypothesized by the ANDREA model.
- Change: According to Thagard and Aubie, "*the key to understanding the onset and cessation of emotions is working memory, ...*". Here working memory is defined as the part of long-term memory, that currently is most active. Working memory, at the neural level is characterized by four crucial aspects: *1*) '*stimulation*', the mechanism by which new information, from external stimuli or other processes, enter working memory and, therefore, consciousness; *2*) '*recurrent activation*', which describes how information, already present and active in working memory, is maintained; *3*) '*decay*', which is the opposite of recurrent activation, and defines the process by which content, that is not actively maintained, slowly fades away; *4*) '*inhibition*', specific to the context of the Emocon model, it describes the process by which activation of some elements, tend to suppress the activity of others. Therefore, within the Emocon model, emotional

experiences begin, when particular neural populations are stimulated, in working memory. The experience continues, if recurrently maintained, until it is either inhibited by another element or simply decays with time.

- Integration and differentiation: Integration, similarly to the representation component, is an inherent feature of the Emocon architecture, which involves so many different brain areas, whose very function is to combine low-level perceptions with high-level inferences. As for differentiation, the authors believe that it can only be achieved through cognitive appraisal. However, as already mentioned, this is the only missing component in the Emocon architecture, although both the EACO and NECO models attempt to account for it.

Even though, Thagard and Aubie (2008) could not implement the complete Emocon architecture, and corroborate its full capabilities, they still managed to provide solid explanations for the different components and emotional phenomena, including somatic representation, emotional decision-making, cognitive appraisal and working memory. Combined, these elements serve to describe a unified theory of emotional consciousness. As a consequence, the authors consider Emocon to have accomplished at least two major contributions: *1*) it has gathered under the same umbrella previously disjoint models and ideas; *2*) and given an account of the central aspects of any emotional experience, such as integration, differentiation, valence, intensity and change.

Although Thagard and Aubie (2008) affirm that the theory related to the Emocon architecture seeks to explain the emergence of emotional experience from the interaction of different cortical areas, it also needs to explore a wide range of functions related to emotional expression. Though, it is interesting to note that to investigate the concept of subjective feelings and their realization in the brain, Thagard and Aubie (2008) first need to find a suitable definition for consciousness. Consequently, it means that for any animal to feel anything it needs some form of consciousness. A similar argument can then be made for robots and virtual agents.

Using the hierarchy introduced by Ortony et al. (2005) in Sub-Section 2.6.3, because the Emocon model describes the emotional mechanisms responsible for subjective feelings and their relationship to different cognitive processes, it is therefore constrained to the routine and reflective levels. According to Ortony et al. (2005), the reflective level is indeed where full-fledged emotions occur, but it is also where emotions have the least influence. The reflective level is the domain of higher-level cognition. Emotions coming up from the routine level are

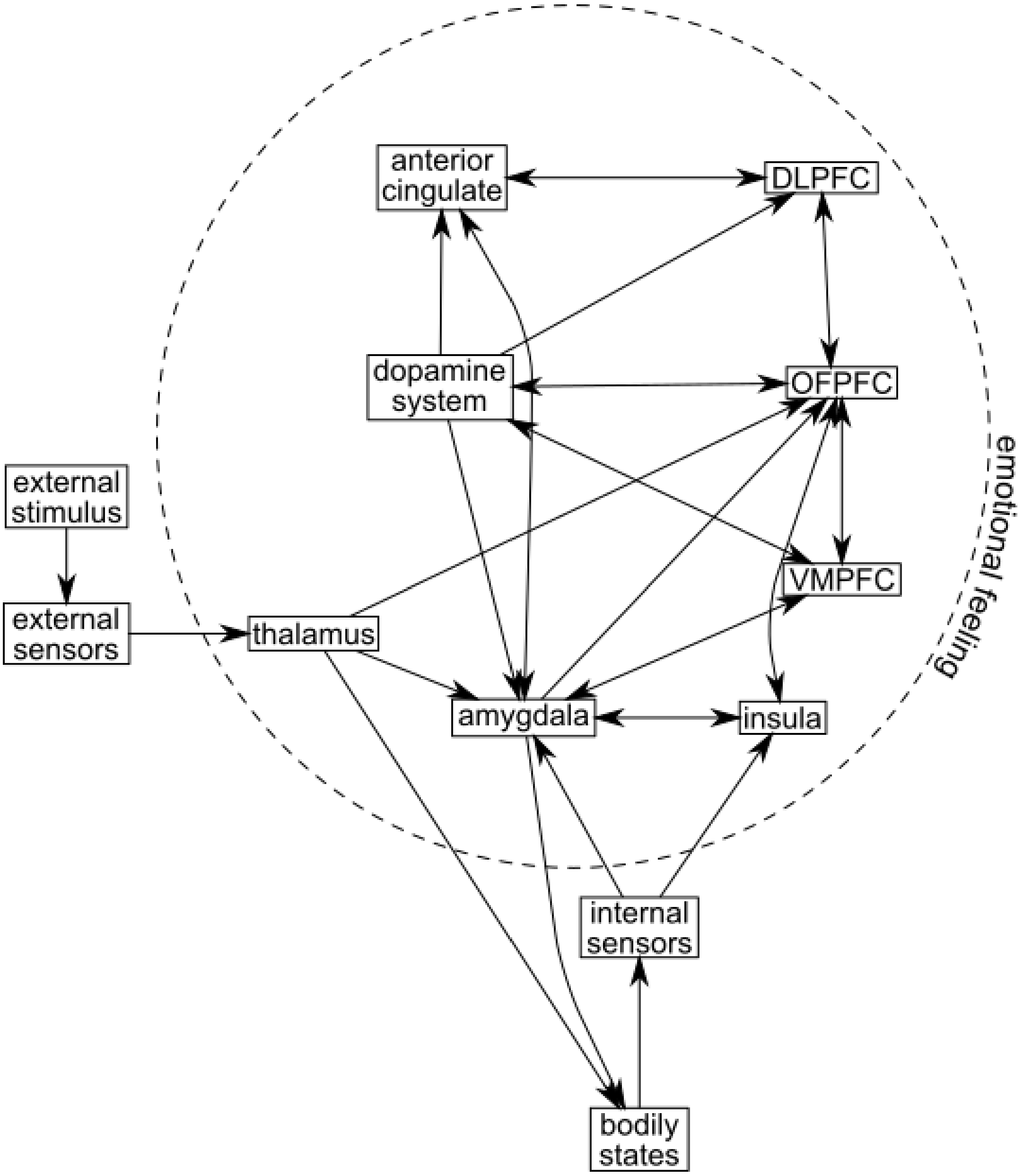


Figure 2.14 *The EMOCON model incorporates concepts from the ANDREA, GAGE and NECO models. Hence, even if it has not been implemented yet, according to Thagard it holds the key to emotional consciousness. It is conjectured that emotion, and therefore by extension emotional consciousness, is not the result of a single brain area's processing. Instead, emotions are thought to emerge from the pattern of neural activity and the interactions of the different systems making up EMOCON.*
Source: Thagard and Aubie (2008, p.817), reproduced with permission from Elsevier.

perceived, contextualized, and represented by a simple label. Even if for many this is where subjective feelings give colors to our lives, since we are finally made aware of their existence, those feelings are only a pale shadow of what emotions were at the reactive and routine levels.

Consequently, even though there is some overlap, the theory developed by Thagard and Aubie (2008) is focused on the cortex and the interactions between it is made of. On the contrary, this thesis stands at the opposite end of the spectrum (as defined by Ortony et al. (2005), as well as Alexandrov and Sams (2005)) investigating the sub-cortical survival circuits and their involvement in the decision-making process. Furthermore, by focusing on emotional expression rather than subjective feeling this thesis avoids having to define consciousness. However, the concept of emotion still needs to be specified.

## 2.7 What definition for emotion?

As made abundantly clear in this chapter answering the seemingly simple question of "*what is emotion?*" is quite complex and a far reaching problem. Without any pretense of completely defining emotion, this section explore the concept of full-fledged emotions and their relationship with cognition. As alluded to in Section 1.5 it is argued that emotion could be construed as an artificial construct emerging from the activation pattern of the cognitive systems engaged to cope with a situation relevant to the individual's survival. However, for the analysis and interpretation of the results in Sections 5.3 and 6.4 to have any meaning a working definition is required. The limitations of such this definition are identified at the end of this section.

### 2.7.1 Full-fledged emotion

Papez (1937) correctly pointed out that what is commonly referred to as '*emotion*' or '*full-fledged emotion*' can be divided into two dissociable concepts: emotional expression and subjective feelings. With the exception of the Emocon architecture (presented in Sub-Section 2.6.4), the theories and models introduced in this chapter all focus primarily on the unconscious processes underlying emotional expression. Functionally, emotional expression is responsible for improving the individual's chances of survival, and, therefore, of furthering the species. Within this context, the emotional system detects stimuli (whether internal or external) of major concern for the organism, arouses, and influences the appropriate cognitive systems whose task it is to ensure that the body is ready to cope with a given situation and/or to bring the individual's internal state back to homeostasis. In addition to monitoring the state of both the environment and the individual, as well as regulating the body's homeostasis, emotions are also used when communicating with conspecifics. If a person is unable to cope with a situation, for example, expressing their helplessness allows them to recruit the help of others in facing the threatening event or object (see Damasio (1997, 2008); Darwin (1872); J.-M. Fellous and Arbib (2005a); LeDoux (1998); Scherer et al. (2010) for more examples of the use of emotions in communication). From a more theoretical perspective, emotional expression is usually defined as "*arising from patterns of individual judgment concerning the relationship between events and an individual's beliefs, desires, and intentions, sometimes referred to as the* person-environment relationship." (Marsella et al., 2010, p. 27). The description of subjective feeling, then, stems from the definition of emotional expression and requires some form of consciousness. As specified by LeDoux (2012), and to some extend by Thagard and Aubie (2008) as well, an individual is only able to experience an emotion whenever their consciousness either detects the activation of a survival related system or perceives the physiological changes and mental states resulting from the emotional system triggering a coping reaction.

### 2.7.2 Emotion as emerging from cognitive activity

As briefly brought up in Section 1.5, a possible interpretation of this thesis' content could be that emotions are pure artificial constructs emerging from the pattern of activity of the cognitive systems recruited to cope with a situation of major concern for the survival of the individual. Similar to how working memory, and to a certain extent consciousness, are theorized to arise from the spontaneous and synchronous activities of different brain systems (Baars, 2005), emotions when considered in this new light would materialize from the activation pattern of the cognitive systems currently engaged in maintaining the body's homeostasis. This conjecture is derived from two main line of thoughts.

Primarily, it can be inferred from the different definitions of emotion introduced in this Chapter. To begin with both Aristotle (350 B.C/1926) and Descartes (1649) defined passion in a fashion very similar to how perception is construed nowadays. Additionally, in his correspondences with Hyperaspistes in August 1641 Descartes openly admitted that he believed the passion and action of the soul to be one and the same. The only difference between the two concepts is one of point of view. While the action of the soul is described from the viewpoint of the individual triggering an event, the soul's passion is an interpretation of the same event from the perspective of the person being passively influence by it. This piece of information could be dismissed on the grounds that the knowledge on which the definition from Aristotle (350 B.C/1926) and Descartes (1649) are based has since been disproved. However, even Arnold (1959) when writing about her theory described appraisal as "*immediate, intuitive and as inevitable as sense perception itself. ... Usually, the sequence perception-appraisal-action is so closely knit that our everyday experience's never the strictly objective knowledge of a thing, ...*" (Arnold, 1959, p. 407). Later, Marsella et al. (2010, p. 27) and Scherer et al. (2010, p. 49) would also describe emotion as either "*arising from patterns of individual judgment ...*" or "*An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems.*". Even J. a. Russell (2003) a dimensional theorist wrote in his paper:

> "*Objects and events all have affective qualities, the perception of which is as much a part of contact with the external world as perception of nonaffective qualities. Indeed, just as sensation, perception, and cognition cannot be cleanly distinguished from on another, perception of affective quality is another part of this processing of information.*" — J. a. Russell (2003, p. 149)

All four authors definitely hinting at the fact that what is referred to as emotion depends on and is hardly distinguishable from cognition.

Second, it is an idea indirectly brought up in LeDoux's (2012) description of full-fledged emotions as being labels attributed to global organismic states when perceived by consciousness (see Sub-Section 2.6.2 for more details). Additionally, in the same article he clearly states that "*In this Perspective I, therefore, describe a way of conceiving phenomena important to the study of emotion, but with minimal recourse to the terms emotion or feelings.*" (LeDoux, 2012, p. 654). As a matter of fact, as will become apparent throughout this thesis, the functions assigned to almost all the brain systems introduced could be construed as purely cognitive in nature.

Consequently, recognizing that the evidences presented are rather weak, it is still possible to conjecture that the distinction between emotion and cognition is an artificial one. Although, it has just been argued that emotion could be considered as a by-product of the activity of the cognitive systems recruited to cope with a survival relevant situation, the remainder of this thesis still separates emotion from cognition (however artificial this differentiation is) for the sake of clarity, and also because this idea has not found unanimous support yet (although it has been discussed in J.-m. Fellous and LeDoux (2005)). Furthermore, in the context of this thesis and for the computational models described within it to be accountable it is necessary to at least provide a working definition for the concept of emotion.

### 2.7.3 A working definition of emotion

Although LeDoux's (2012) '*survival circuit*' theory does not suggest any new concept, it borrows ideas introduced throughout history and across affective science's many sub-fields, from Darwin (1872) to MacLean (1970) and from appraisal to anatomic theories, and rearranges them into a coherent narrative that emphasizes the relationship between survival and emotions. To a certain extent, in my opinion, LeDoux's (2012) reinterpretation of those concepts appears to be the closest in following in Darwin's (1872) footsteps. Furthermore, the neural structure he describes as being essential for the survival of animals has also been identified and analyzed in details in other publications (Panksepp (1998), Damasio (2008), Whalen and Phelps (2009) and Niv (2009), to name but a few). Thus, it could be considered as standard within the neuroscience community. Consequently, it should come as no surprise that the remainder of this thesis takes inspiration from the '*survival circuit*' theory.

As a result, the definition of emotion adopted for the rest of this thesis is similar to the description provided by Scherer et al. (2010) and LeDoux (2012). Emotion, therefore, is construed as a mental state emerging from the activation of the cognitive systems recruited by the brain to cope with a situation of major concern to the individual. If emotions are to be considered as mental states, however, it means that we also need to define features specific to emotions allowing us to differentiate them from other mental states. Across the different theories and models introduced in this chapter, it seems that emotions are recurrently associated with a '*feeling*' of pleasure or pain. Moreover, their effect on the brain and the individual in general is characterized as being either exciting or depressing. Consequently, in the context of this thesis emotion is differentiated from other mental states by its level of arousal and valence. Although, J. a. Russell (2003) mentions in his article that those two features are not enough to account for full-fledged emotions, because the project described here is only interested in the primitive form of emotion the definition presented here should suffice.

### 2.7.4 Limits of defining emotion

As illustrated by the fictional discussion between Russell and Edison (J.-M. Fellous & Arbib, 2005b), providing a complete definition of full-fledged emotion is a daunting task as it involves a cascade of other definitions, such as consciousness or fear. Given that this exercise is already arduous for humans and animals, one can only imagine the difficulty of describing emotion for artificial agents or robots, for which even the meaning of pain has to be reinvented. From the debate between Russell and Edison, it can be concluded that any serious attempt at fully defining emotion is either futile, since the number of concepts to on which emotion depends grow exponentially, or ultimately results in a flawed definition influenced by the individual's perspective. Still, Russell argues that even an incomplete definition if specific enough can be of use, since it allows "*meaningful comparison between different approaches to humans, animals, and machines*" (J.-M. Fellous & Arbib, 2005b, p. 5).

Following Russell's argument, the definition of emotion adopted for this thesis is appropriate, however restrained it might seem. Indeed, as discussed in Chapter 1 this thesis' focus is on emotional expression, or more specifically on the primitive form of emotion and its role in survival. Consequently, using a characterization that accounts for full-fledged emotion would unnecessarily complicate the issue and possibly overshadow some of the finer differences between the computational models introduced and other similar architectures. Therefore, it is my opinion that conceiving of emotions as a mental states emerging from the activity of the cognitive systems in charge of maintaining the body's homeostasis, and differentiated by their level of arousal and valence is sufficient to deliberate about the overall hypothesis.

# Summary

Although the concept of emotion has attracted the attention of many philosophers and scientists, from ancient Greece to today, it does not have an agreed upon definition yet. As a result of the multi-faceted nature of emotions, especially in humans, many theories have been developed over time to try and answer the question 'What is Emotion?' Thus, this chapter explores the point of view adopted by three sub-fields of affective science. Within the dimensional theory, 'affect' is represented as a point in a dimensional space, 'pushed' around by the 'affective quality' of the objects a person interacts with. Therefore, dimensional theories focus on defining the dimensions required to fully describe affect. In the context of appraisal theory, emotions are considered to be simple labels attributed to stimuli by different cognitive systems. Hence, the goal of this sub-field is to detail the processes used by cognition to determine the relevance of a given stimulus for an individual's goals, beliefs and desires. The last sub-field resides at the intersection between affective science and neuroscience, as such it seeks to understand the neural underpinnings of emotions in the animals brain. Based on the principle of the 'survival of the fittest', Darwin (1872) theorized that emotions evolved as a mechanism to increase the survival potential of animals. Consequently, there is a set of emotions that is shared not only across cultures, but also across species. Furthermore, the brain structures involved in processing these emotions should be conserved between species. Following in Darwin's (1872) footstep, LeDoux (2012) introduces the 'Survival Circuit' theory, in which he describes emotions as being labels assigned by the conscious mind to 'global organismic' states, whenever it detects the triggering of innate circuits responsible for maintaining the body's homeostasis. Given that the neural structure identified by LeDoux (2012) are though to improve the survival potential of animals, the remainder of this thesis will take inspiration from the 'survival circuit' theory. As a result, and in spite of the fact that emotions have been suggested to be a by-product of the activity of the cognitive systems in charge of the body's homeostasis, a definition of emotion similar to those suggested by LeDoux (2012) and Scherer et al. (2010) has been adopted for the project described here after. Therefore, emotion is conceived as a mental state resulting from the activation of the cognitive systems recruited by the brain to cope with a situation of major concern for the individual. Finally, emotional states are differentiated from other patterns of brain activity by their level of arousal and valence. Two features that seem to be recurrently attributed to emotion.

CHAPTER 3

# Decision making

As established in Chapter 2 two properties used to characterize emotions seem to be recurrent. They are best expressed using the terms ‘*valence*’ and ‘*arousal*’ borrowed from the dimensional theory. Valence within the brain can take several appearances. However, when considered in terms of the functionality emotions fulfill, it usually translates to ‘*reward*’ and ‘*punishment*’. This chapter goes further and explores how emotions, especially reward and punishment, are used by the brain to manage its limited resources (both motor and cognitive). It describes the path from classical conditioning to reinforcement learning, then moves on to the dopamine and serotonin systems, identified by neuroscientists as the main areas signaling reward and punishment in the brain. Finally, the actor-critic model is explored in more details, from its mathematical roots to its computational model, then to its implementation inside the brain. Along the journey we will strive to answer the question: how does the brain learn and decide what to do in any given situation?

## 3.1 Reward and punishment in the brain

Although reinforcement learning is nowadays well-known as a flourishing machine learning paradigm, it evolved from the need to model results from classical conditioning. First introduced by R. A. Rescorla and Wagner (1972), as the delta-rule, it waxed and waned before being brought back to life thanks to the work of Sutton and Barto (1998). Their ideas helped build the foundations for the models developed today. Furthermore, it spiked the interest of neuroscientists who again used those same equations to interpret empirical data on classical conditioning. Hence, lending more support to the theory and guiding the search for the brain mechanisms underlying conditioning.

### 3.1.1 The dopaminergic system

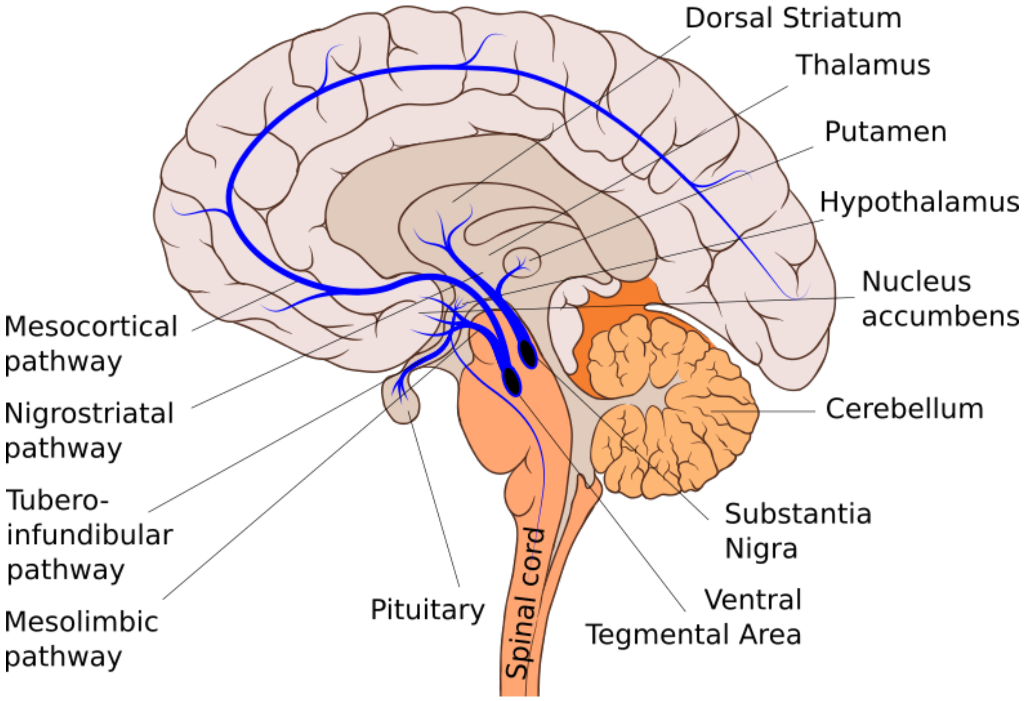


Figure 3.1 *A representation of the dopamine (DA) system's influence over the human brain. Dopamine signaling originates from two main areas, namely the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc).*
Source: commons.wikimedia.org, Authors: Slashme, Patrick J. Lynch and Fvasconcellos.

For a long time the dopaminergic system has been split between two roles. On the one hand, there are scientists who attributed an excitatory role to the dopamine neurons. And on the other hand there were those, who thought that dopamine was some kind of a reward signal (Cools, Nakamura, & Daw, 2011). It turns out that history and biology has proved both parties to be right to some extent (Niv, Daw, Joel, & Dayan, 2007).

The dopaminergic system is made of two main areas (as shown in Figure 3.1) the ventral tegmental area and the substantia nigra pars compacta (SNc) (Seamans, 2007). These two parts of the midbrain have two levels of functioning. The first one, called '*tonic*', can be associated with an average firing rate. It constitutes some sort of baseline and maintain a minimal amount of dopamine in the brain at all time. The second one, called '*phasic*', is defined by burst of dopamine. That is a sudden increase in the firing rate followed by a rapid decrease to finally match the tonic level again (Niv, 2009).

It is because of those two modes of functioning that researchers have had a hard time agreeing on the exact role of the dopamine system. However, recently Niv, Daw, Joel, and Dayan (2007) suggested a model to try and unify all the experimental results gathered so far. Thus, in their model the tonic level of dopamine has been associated with an excitatory effect on the motor area, leading to more vigorous movements or an increased '*motivation*', while the phasic bursts were defined, through the reinforcement learning framework (Arbib, 2003; Sutton & Barto, 1998), as reward prediction error. It is worth mentioning that in reinforcement learning, even if the focus is on phasic dopamine levels, the tonic level is also used as the baseline or limit between reward and punishment (a burst over the tonic level being a reward and a dip under defining a punishment).

### 3.1.2 The serotonergic system

Despite having an overall much greater influence on the brain than the dopaminergic system (see Figures 3.1 and 3.2) and five times its number of receptor's sub-types, hinting at a very diversified role the serotonergic system does not have an explanatory framework comparable to reinforcement learning (Cools et al., 2011).

Based on the experimental results obtained through animal instrumental conditioning, but also from patients suffering from mental disorders, and on the fact that the serotonergic and dopaminergic systems are supposed to have complementary roles, Cools and colleagues hypothesized that the serotonergic system is involved in slow and careful movements, lack of motivation to act and the perception of punishment (Cools et al., 2011). Lastly, as depicted in Figure 3.2, the serotonergic system completely depends on the raphe nuclei, especially the caudal and rostral part. More interesting, however, is the fact that activity from the serotonin neurons are inhibiting the dopamine neurons, whereas the dopaminergic system has no direct control over the serotonergic system (see Figure 3.1 and Cools, Nakamura, and Daw (2011); Daw, Kakade, and Dayan (2002); Litt, Eliasmith, and Thagard (2006)).

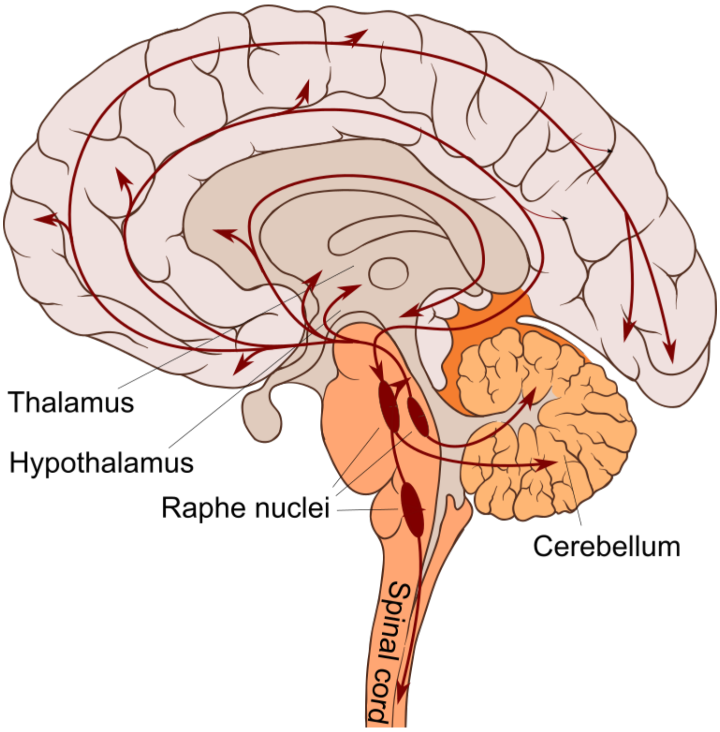


Figure 3.2 *This is a depiction of the serotonin (5-HT) system. Signals mainly originate in the raphe nuclei, especially the dorsal portion. Compared to the dopamine system, it has a wider influence both within the cortex as well as in sub-cortical areas. More importantly, though, the serotonin system has projections to the VTA and SNc. Projections that are not reciprocated by the dopamine system.*
Source: commons.wikimedia.org, Author: Sven Jähnichen.

### 3.1.3 Synergistic interactions between dopamine and serotonin

As mentioned previously, the roles of the dopaminergic and serotonergic systems are thought to be complementary (Cools et al., 2011). Thus, while dopamine has an excitatory effect on neurons, serotonin would be inhibitory.

However, quite recently Di Pietro and Seamans (2011) have made quite a surprising discovery. By subjecting pyramidal cells with dopamine alone, then serotonin alone and finally different combinations of dopamine and serotonin, at different concentrations, they observed that both dopamine and serotonin alone have an excitatory effect on the cells. Mixing dopamine and serotonin, with a higher concentration of dopamine, resulted in an even higher excitability in the pyramidal cells. On the contrary, a mix of dopamine and serotonin, with higher concentration of serotonin would inhibit pyramidal neurons' activity. This fact would then give another explanation for the existence of a tonic level of dopamine. Indeed, if serotonin is to play the role of the inhibitor, then it will need to mix with some dopamine or else the desired effect could never be achieved.

The study of the dopaminergic system has known a sudden increase in interest after the introduction of the reinforcement learning framework (Arbib, 2003; Sutton & Barto, 1998), as it is considered to be a mechanism for reward prediction error. However, as detailed in this section and put forth by other authors, it is clear that the dopaminergic system alone cannot account for every aspects of reward and punishment in the brain (A. K. Anderson et al., 2003; Cools et al., 2011; den Ouden et al., 2013; Di Pietro & Seamans, 2011; Yates, Darna, Gipson, Dwoskin, & Bardo, 2015). Hence, efforts should be made to reveal the remaining mysteries surrounding the serotonergic system.

## 3.2 Reinforcement learning

Born from experiments using classical and instrumental conditioning (two procedures defined in the first Sub-Section), the field of reinforcement learning has steadily developed to become an inescapable part of today's machine learning community. Before analyzing one of the standard model of reinforcement learning that was developed in the early 1990s and known as the actor-critic model, this section explores the general concepts that form the core of the reinforcement learning framework. Finally, the link between reinforcement learning and the animal's brain (human included) is detailed.

### 3.2.1 A word about conditioning

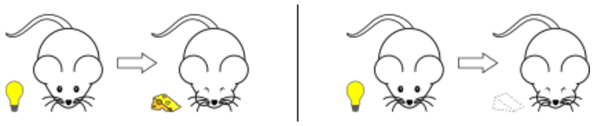


Figure 3.3 *This figure depicts the two stages of a typical classical conditioning experiment. In the first stage, the mouse is presented with both the conditioned (CS) and unconditioned (US) stimuli simultaneously. During this stage, it learns the association between the two stimuli. For the second stage, only the CS is presented, which triggers the innate response associated with the US.*
Source: commons.wikimedia.org, Author: Nicolas P. Rougier.

Most, if not all, the experiments aiming at observing the effects of reward and punishment in the brain are based on some form of conditioning. At the present there exist two main classes of conditioning, namely classical conditioning, discovered by Pavlov (1927) and later extended by R. A. Rescorla and Wagner (1972), and instrumental conditioning pioneered by Thorndike (1911).

In the realm of classical conditioning experiments, first described by (Pavlov, 1927), animals are subjected to both a stimulus with no particular rewarding or punishing value, called the '*Conditioned Stimulus*' (CS), and an innately rewarding or punishing stimulus, like a foot shock or food, called the '*Unconditioned Stimulus*' (US). After repeatedly presenting simultaneously the CS and US to the animal the CS will become associated with the rewarding or punishing value of the US. So that in the end the CS alone will trigger an innate response from the animal. In his well known experiment, Pavlov subjected his dogs to a particular tone before feeding them. After some short period of time, playing the tone alone was enough to make the dogs salivate (Pavlov, 1927).

In contrast, instrumental conditioning involves the subject learning how to perform a task, the goal of which is signaled by the delivery of a reward or punishment (US). The standard instrumental conditioning experiment would involve an animal sitting or placed in front of two lights. Whenever a light is switched on the animal will then have, after a certain delay, to push a button or pull a lever placed under each light. If the animal performs the right action a reward will be delivered in the form of food pellets or fruit juice (US). However, if the animal chooses the wrong button, nothing happens, which can be perceived as a punishment or lack of reward, in this case.

In the point of view of the observer, instrumental conditioning has two main advantages over classical conditioning. First, for classical conditioning to work properly the CS and US should be presented simultaneously or with a delay no longer than one or two seconds. Otherwise, the stimuli will not become associated. Second, because in instrumental conditioning the animal has to perform an action to get a reward, the animal's reaction time can be used as a direct measure of its motivation.

As will be detailed later in this section these two advantages are the reasons for scientists to prefer instrumental conditioning to classical conditioning, when performing experiments on reward and punishment.

### 3.2.2 Characteristic features

The earliest trace of reinforcement learning in history goes back to Thorndike (1911), already mentioned for his contribution to instrumental conditioning. It is not surprising to see that Thorndike's work was the basis for both instrumental conditioning and reinforcement learning, as reinforcement models were first developed to account for results from experiments on reward perception in animals and humans (Dayan & Abbott, 2001; Sutton & Barto, 1998). As part of his pioneering work Thorndike introduced the '*law of effect*'. Hence laying down the foundation for what would become reinforcement learning as it is known today. At an abstract level, the law of effect describes the learning process at the core of this paradigm:

> "*Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.*"
> — Thorndike (1911, p. 244)

Later pioneers such as Rescorla and Wagner would be responsible for bringing reinforcement learning to the fore. However, researchers would quickly come to realize that the original equations at the core of the framework were ultimately flawed. Furthermore, the rise to prominence of supervised learning and artificial neural network spelled the downfall of the field. In the literature this event is usually referred to as the '*first fall*' of reinforcement learning (Sutton & Barto, 1998). Its revival was achieved through two different paths. The first one pioneered by Bellman (1957) considered reinforcement learning as a way to achieve optimal control and lead to the introduction of dynamic programming. The second path, which is most commonly associated with the idea of reinforcement learning, is that of trial-and-error learning (Sutton & Barto, 1998). Because the main focus of reinforcement learning nowadays is on trial-and-error, the remainder of this section will develop this second concept in more details.

Reinforcement learning, contrary to popular belief, is not a unique algorithm or model, but a framework describing how an agent can learn to accomplish a task through interactions with its environment. Consequently, any model or algorithm capable of answering the '*reinforcement learning problem*' is categorized as a reinforcement learning algorithm. The problem, at the core of the framework, is to find the optimal policy for completing a given task. The optimality parameter being measured by the total amount of reward an agent perceives over the total amount of time it takes to complete the task. To solve this problem most of the standard algorithms have at their disposal four elements (Sutton & Barto, 1998):

1. A value function: Depending on the algorithm, each state or state-action pair is associated with a value representing the total amount of reward an agent perceived when reaching the state $s$ or when reaching the state $s$ by executing action $a$. The value function is usually denoted $V_\pi(s)$ or $V_\pi(s,a)$ (note that the value function depends on the chosen policy $\pi$).

2. A policy: Denoted by the letter $\pi$. The policy describes the agent's strategy for choosing the next action to perform, when a decision is required.

3. A reward function: It maps each state of the environment with a reward $r$, that indicates the desirability of the state. The reward function is also used to indicate to the agent the goal of the task. Thus, in a maze task for example, the reward for all the states would be 0 except for the state in which the agent has solved the maze, in this case the reward would be set to 1.

4. And optionally a model of the environment: Depending on the type of algorithm a simplified representation of the environment, a model, can be provided to the agent, allowing it to make predictions as to the consequences of its actions. Thus, it might serve as a way to optimize the search strategy. The presence of this element leads to a crude categorization of reinforcement learning algorithms into two categories: model-based, inherited from the control theory path, and model-free relying solely on the interactions with the environment.

### 3.2.3 The Rescorla-Wagner equation

In 1972, based on the results of animal classical conditioning experiments, Rescorla and Wagner (R. Rescorla, 2008; R. A. Rescorla & Wagner, 1972) hypothesized that, animals would update the expected reward associated with a compound conditioned stimulus (CS), based on the discrepancy between the current expectation and the value of the unconditioned stimulus (US) presented simultaneously. Consequently, assuming that the agent's state, during a conditioning trial, is described by a real-valued vector of stimuli: $\mathbf{u}(S) = (u_1(S), u_2(S), \ldots, u_n(S))^\top$, where $u_i(S) = 1$ denotes the presence of $CS_i$ ($i^{th}$ component of the compound CS), and $u_i(S) = 0$ its absence, then the '*value estimate*' of state $S$ can be expressed as:

$$v(S;\omega) = \omega^\top \mathbf{u}(S) \tag{3.1}$$

Furthermore, if $t$ indexes each conditioning trial, then the vector of associative strengths ($\omega_t$) learns between trials using the equation:

$$\omega_{t+1} = \omega_t + \varepsilon \times \delta_t \times \mathbf{u}(S_t) \tag{3.2}$$

Where $\varepsilon$ is the learning rate and the '*prediction error*' ($\delta_t$) reflects the difference between the magnitude of the unconditioned stimulus and the state's estimated value:

$$\delta_t = r_t - v(S_t;\omega_t) \tag{3.3}$$

Even though the Rescorla-Wagner model was at the basis of the revival of the reinforcement learning framework, in spite of its power of prediction regarding results from classical conditioning experiments, the model fell short when trying to explain the concept of second order conditioning (Dayan & Abbott, 2001; Niv, 2009; Sutton & Barto, 1998).

Second order conditioning is based on the same concept than classical conditioning. It can be described in two main steps. First, as for classical conditioning, the subject will be simultaneously presented with a conditioned stimulus ($\mathrm{CS}_1$) and an unconditioned stimulus (US). This stimulation will be repeated until $\mathrm{CS}_1$ is associated with the unconditioned response (or innate response, UR). Then as a second step, the animal will be subjected to $\mathrm{CS}_1$ and to a second $\mathrm{CS}_2$. In the same way as before, after repeated exposure $\mathrm{CS}_2$ will be associated with the UR. Thus, presenting either $\mathrm{CS}_1$ or $\mathrm{CS}_2$ will have the animal react using the associated UR (Schmajuk, 2008).

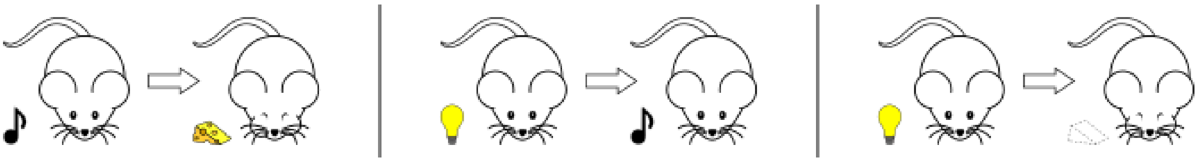


Figure 3.4 *This figure depicts the three stages of a typical second order conditioning experiment. In the first stage, the animal is simultaneously presented with both unconditioned (US) and conditioned (CS1) stimuli. Its brain learns to expect the association between the two stimuli. For the second stage, the animal is subjected to a new conditioned stimulus (CS2), as well as CS1. Since CS1 can already trigger an innate response, the animal learns to associate CS2 with CS1. Finally, in the last stage, presenting CS2 to the animal should trigger the innate response associated with the US.*
Source: commons.wikimedia.org, Author: Nicolas P. Rougier.

### 3.2.4 Temporal difference learning

To account for the introduction of the time parameter in the conditioning experiment, Sutton and Barto (1998), developed the well-known Temporal Difference learning algorithm, usually referred to as TD-learning. Borrowing ideas from the Monte Carlo methods as well as dynamic programming, one could identify the TD-learning algorithm as the fresh new core of the reinforcement learning framework.

Considering the Rescorla-Wagner model again (see Equation 3.2), it appears that the update mechanism is too short sighted, since it accounts only for immediate reward (Niv, 2009; Sutton & Barto, 1998). However, for second order conditioning and sequence learning in general (as is the case for instrumental conditioning), the agent must be able to plan more than one step ahead, sometimes without knowing the consequences of its actions until after reaching the goal. The insight, therefore, was to introduce an '*eligibility trace*', as well as a '*discount factor*'. The eligibility trace, looking backward, allows previously visited states to influence to some degree how the value of the current state is updated. On the contrary, the discount factor looks forward and allows future outcomes to influence the updating process as well. It also accounts for the fact that animals, in general, prefer immediate rewards to those in the distant future. If $\mathbf{z}_t$ is a vector of eligibility traces, then the update expression used in the Rescorla-Wagner model becomes:

$$\omega_{t+1} = \omega_t + \varepsilon \times \delta_t \times \mathbf{z}_t \tag{3.4}$$

Hence, replacing the vector of features $\mathbf{u}(S_t)$, describing a given state $S_t$, with the vector of eligibility traces. Furthermore, it should be noted that in the case of TD-learning, $t$ instead of indexing complete conditioning trials, denotes a single time step.

Similarly, the prediction error $\delta_t$ introduced in the Rescorla-Wagner model is transformed into the following TD error:

$$\delta_t = r_{t+1} + \gamma \times v(S_{t+1}; \omega_t) - v(S_t; \omega_t) \tag{3.5}$$

Finally, the components of the vector of eligibility traces are incremented or decremented according to the stimuli present in $\mathbf{u}(S_t)$ or are slowly decayed, at a rate $\lambda$ (usually referred to as the eligibility trace decay parameter), following the expression:

$$\mathbf{z}_{t+1} = \gamma \times \lambda \times \mathbf{z}_t + \mathbf{u}(S_t) \tag{3.6}$$

Since its revival in the 1970s, reinforcement learning has flourished and has found a plethora of applications in the industry as well as the scientific community. In particular two algorithms, namely the SARSA and the Q-learning (Watkins, 1989) algorithms, seem to have established themselves at the fore of the field of reinforcement learning (Woergoetter & Porr, 2008).

## 3.3 The Actor-Critic

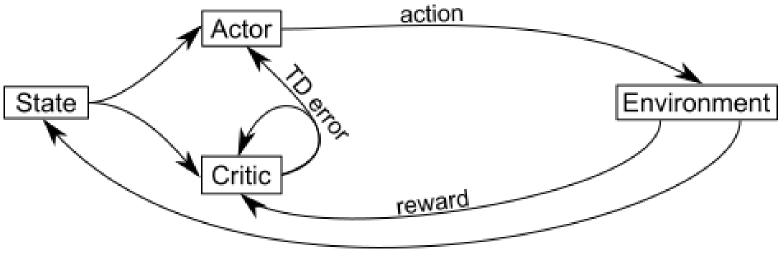


Figure 3.5 *The Actor-Critic model, introduced by Sutton, is a reinforcement learning algorithm. It is made of two components, namely the actor and the critic. The actor, as the name implies is responsible for optimizing the agent's policy. Hence, it takes a state or state-action pair as input and outputs the next action to perform. The critic, on the other hand, is responsible for learning the value of each state or state-action pair. As a result, based on the same input as the actor in addition to the actual reward given by the environment, the critic computes the error in reward prediction, also called the* 'TD-error'. *The critic's output serves as a learning signal for both components of the model. On the contrary to other TD-learning algorithms, that only focus on learning the value function using a fixed policy, the Actor-Critic model optimizes both at the same time.*

Having described the main features of the reinforcement learning framework, this section introduces the '*Actor-Critic*' model. As a reinforcement learning method, this model has recently grown in popularity, all the while being the precursor to the other reinforcement learning algorithms mentioned above, such as SARSA and Q-learning (Barto, Sutton, & Anderson, 1983; Sutton, 1984; Sutton & Barto, 1998). In addition to being a widely used algorithm within the machine learning community the actor-critic has also gathered the interest of many neuroscientists. As is the case for the temporal difference learning algorithm, which is regarded as a model for dopamine signaling within the animal's brain, so too have scientists tried to use the actor-critic model to describe how the brain solves the problem of '*action selection*' or '*decision-making*'.

### 3.3.1 Mathematical model

The reinforcement learning framework described so far provides a useful model to interpret data from classical conditioning experiments, but is limited in its ability to explain results from studies on instrumental conditioning (Niv, 2009). In classical conditioning experiments, regardless of how the subject reacts to the conditioned stimulus, he will always end up in the same state and obtain the same reward. After hearing the telltale tone, whether the mouse tries to flee or stays put or even cower in fear, in the end it will get a foot shock. Consequently, the subject only learns to predict future reward or punishment associated with different stimuli. On the contrary, in instrumental conditioning, the delivery of a reward is contingent on the subject performing the correct action. There is therefore an incentive for the subject to optimize its decision-making strategy to increase the total amount of reward it can get. The problem of optimal action selection is compounded by the fact that in complex and dynamic environments, the reward for accomplishing a given task is delayed until after its completion. Therefore, one has to look back after the facts to determine which were the key actions that lead to the realization of said task. This is known as the '*credit assignment*' problem (Barto et al., 1983; Niv, 2009; Sutton, 1984; Sutton & Barto, 1998). TD-learning methods solve this problem by using the immediate outcome, as well as the expected reward for each state to decide what action to perform next. A straight forward approach, employed in very simple cases, is to greedily select the action which leads to the state with the highest expected outcome.

To provide an elegant answer to both credit assignment and optimal selection problems in a purely model-free environment, Sutton (1984) devised a simple architecture based on two neural populations: the actor-critic. His insight was to realize that, in the absence of any immediate outcome (as is the case for delayed reward), the nature of an action can still be quantified via the discrepancy between the old and the new states (denoted $v(S_t;\omega_t)$ and $v(S_{t+1};\omega_t)$, respectively). Therefore in Equation 3.5, if the reward is set to 0 ($r_t = 0$) the resulting TD error takes the shape:

$$\delta_t = \gamma \times v(S_{t+1};\omega_t) - v(S_t;\omega_t) \tag{3.7}$$

As a result, an action leading to a state with higher value ($v(S_t;\omega_t) < v(S_{t+1};\omega_t)$) will be reinforced, whereas an agent tends to avoid actions, which bring it to a state of lower value ($v(S_t;\omega_t) > v(S_{t+1};\omega_t)$). Consequently, it is possible for the agent to adapt and optimize its policy.

True to its name, the role of the Critic is to critique the action chosen by the Actor. To learn the value of each state, the Critic component uses a standard TD method. Therefore, if $\mathbf{u}(S_t)$ is again the vector of features, completely describing the state $S_t$, and $\mathbf{z}_t^{\omega}$ is the vector of eligibility traces associated with the Critic, then learning within this component is characterized by the following equations:

$$\delta_t = r_{t+1} + \gamma \times v(S_{t+1};\omega_t) - v(S_t;\omega_t) \tag{3.8a}$$

$$\mathbf{z}_t^{\omega} = \lambda^{\omega} \times \mathbf{z}_{t-1}^{\omega} + \mathbf{u}(S_t) \tag{3.8b}$$

$$\omega_{t+1} = \omega_t + \varepsilon \times \delta_t \times \mathbf{z}_t^{\omega} \tag{3.8c}$$

It should be noted that $\delta_t$ the TD error is the same as in Equation 3.5. Moreover, the prediction error is the Critic's only output, which is used by both components to learn.

While the Critic tries to reduce the discrepancy between the current state's value and the immediate and expected outcome, the Actor adapts the policy to keep this same discrepancy positive, since incorrect actions are those that lead to lower value states and therefore are associated with negative TD error. Using the Critic's output, the Actor component updates the agent's policy (denoted $\pi(a_t|S_t;\theta)$, where $\theta$ is the vector of parameters associated with the policy $\pi$) according to:

$$\mathbf{z}_t^{\theta} = \lambda^{\theta} \times \mathbf{z}_{t-1}^{\theta} + \nabla \log(\pi(a_t|S_t;\theta_t)) \tag{3.9a}$$

$$\theta_{t+1} = \theta_t + \varepsilon \times \delta_t \times \mathbf{z}_t^{\theta} \tag{3.9b}$$

The main output of the Actor component is $\theta$ the vector of parameters used to configure the policy and choose the next action to perform. An example of a standard policy, if the action space is discrete and small enough, is to attribute the highest probability of being selected to the action with the highest preference, using an exponential soft-max distribution:

$$\pi(a|S;\theta) = \frac{\exp^{h(S,a;\theta)}}{\sum_b \exp^{h(S,a;\theta)}} \quad \text{defines the policy} \tag{3.10a}$$

$$h(S,a;\theta) = \theta^\top \mathbf{u}(S,a) \quad \text{defines the preferences} \tag{3.10b}$$

### 3.3.2 Decision making in the human brain

In spite of its elegance and popularity, the TD-learning algorithm (and by extension the actor-critic) cannot easily be mapped directly onto brain areas. Furthermore, its reliance on temporal chaining requires that both the timing and sequence from conditioned stimulus (CS) to unconditioned stimulus (US) be exact every time. Hence, making it unsuitable as an explanation for animal learning in instrumental conditioning (Hazy et al., 2010; O'Reilly et al., 2007). Consequently, in an effort to explain the empirical data from both classical and instrumental conditioning, neuroscientists have sought to understand the mechanisms implemented by the animal's brain, that produce those behavioral results (Frank et al., 2001; Grillner, Robertson, & Stephenson-Jones, 2013; Gurney, Prescott, & Redgrave, 2001a; Hazy et al., 2010; Houk, 2011; O'Reilly et al., 2007; Prescott, Gurney, Montes-Gonzalez, Humphries, & Redgrave, 2002; Redgrave, Prescott, & Gurney, 1999; Vitay & Hamker, 2010). Not only has this modeling effort broadened our understanding of the role played by the different sub-cortical areas and their interactions. It has also led to the introduction of both the Primary Value Learned Value (PVLV) architecture and a dual-pathways gating mechanism as a standard model for the basal ganglia.

#### Primary Value Learned Value: a critic within the brain

That dopamine activity is the primary signal for learning within both classical and instrumental conditioning paradigms, is nowadays a widely accepted idea (Berke, 2018; Hazy et al., 2010; Maia, 2009). In spite of this fact, the mechanisms responsible for the phasic dopamine bursts signaling a reward, as well as the influence of dopamine over the learning process of these mechanisms, are not very well understood yet (Hazy et al., 2010; O'Reilly et al., 2007). O'Reilly et al. (2007) attempted to provide an explanation embodied by the '*Primary Value Learned Value*' architecture. Adopting a top-down approach, they first developed a computational model to account for the behavioral data from conditioning experiments. Since a wealth of empirical evidences seemed to indicate that the animal's brain is using mechanisms similar to those hypothesized by the PVLV, its components can easily be mapped to different brain areas. Therefore, providing simultaneously a computational model and a biologically plausible explanation for the brain's mechanisms supporting dopamine activity. The PVLV system being part of an ongoing research project, it has recently been extended, by Hazy et al. (2010), to include a novelty detection component.

The PVLV architecture can be further divided into three systems, namely the PV and LV, which were introduced in the original version, and finally the NV, the most recent addition. The PV and LV systems both have excitatory and inhibitory components. The neural activity of the excitatory component of the PV system ($\mathrm{PV_e}$) represents the primary reward. Its inhibitory component ($\mathrm{PV_i}$) learns to expect the occurrence of a given US. Once training is complete, it is able to inhibit $\mathrm{PV_e}$'s activity. Hence, avoiding any further phasic dopamine burst at US-onset. The excitatory ($\mathrm{LV_e}$) and inhibitory ($\mathrm{LV_i}$) components of the LV system have a similar interaction, the only difference being that they learn to drive dopamine activity for any CS reliably paired with a US. According to O'Reilly et al. (2007) the $\mathrm{LV_e}$ component is capable of associating a CS with a US, because it can only learn to trigger the dopamine system at CS-onset, if a reward is received or highly expected at US-onset. This either requires for the CS and US to co-occur, or the brain to keep the CS in memory until US-onset. Having $\mathrm{LV_e}$'s learning be contingent on the delivery of a reward, avoids any positive reinforcement from self-generated dopamine activity (Hazy et al., 2010; O'Reilly et al., 2007). $\mathrm{LV_i}$ follows the exact same learning process albeit with a slower learning rate. Consequently, it slowly learns to inhibit $\mathrm{LV_e}$'s excitatory effects and implement a habituation mechanism. Finally, the NV component, later added by Hazy et al. (2010), has the single task of driving phasic dopamine activity upon detection of a novel stimulus. Furthermore, because of its learned decay mechanism, repeating stimuli will trigger incrementally lower bursts of dopamine.

Both PV and LV systems share the same learning mechanism, whose core is based on the same learning principle as the Rescorla-Wagner rule (see 3.2.3). This means that it learns from the discrepancy between the predicted and actual outcomes. As detailed above, if the excitatory component of the PV system represents the actual reward ($PV_e = r_t$), and the inhibitory component is the expected outcome ($PV_i = v(S_t, \omega_t)$), then the basic Rescorla-Wagner equation

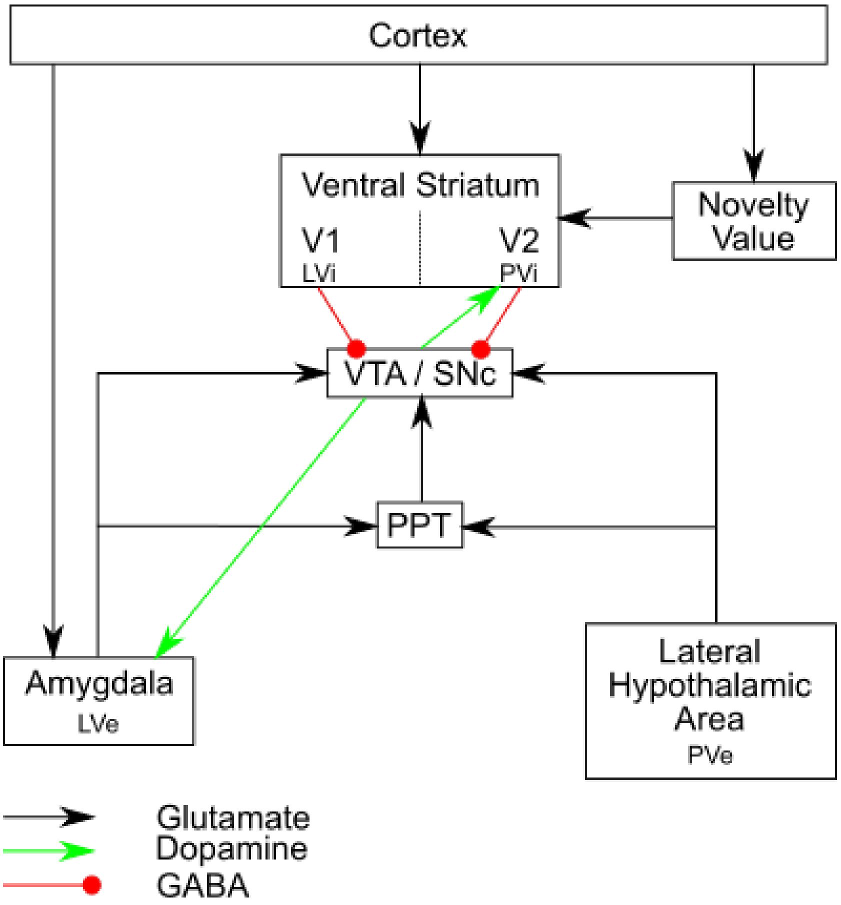


Figure 3.6 *The Primary Value Learned Value (PVLV) model, as introduced by O'Reilly, Is made up of two systems, which can each be sub-divided into two modules. The excitatory component of the PV system (the* $PV_e$*) is triggered whenever a primary reward is received. Its inhibitory component (the* $LV_i$*), on the other hand, slowly learns to expect the occurrence of a given unconditioned stimulus (US). Hence, it implements a habituation mechanism. The LV system, also made of two sub-components, only learns whenever the dopamine system is triggered. The excitatory (*$LV_e$*) and inhibitory (*$LV_i$*) modules have a similar interaction. This means that, while* $LV_e$ *learns to expect a given conditioned stimulus (CS) reliably paired with a US, the* $LV_i$*, slowly suppresses* $LV_e$*'s activity. In this, figure is also present an extension, later suggested by Hazy, that of the Novelty Value (NV) system. This system implements yet another habituation mechanism, by which new unforeseen stimuli trigger a spike in activity, repeating stimuli do so with increasingly less intensity.*
Source: O'Reilly et al. (2007, p.42), reproduced with permission from the American Psychological Association, Inc.

(Equation 3.1) can be re-written as:

$$\delta_{PV} = PV_e - PV_i \tag{3.11}$$

This value is then used to train $\mathrm{PV_i}$ to expect the occurrence of a given US, using the same formula as described in Equation 3.2.

Before detailing the learning algorithm for the LV system, Hazy et al. (2010) define the condition upon which the LV system relies to determine when to learn:

$$PV_{filter} = PV_e \text{ has a value or } (PV_i > 0.8 \text{ or } PV_i < 0.2) \tag{3.12}$$

It should be noted that $PV_{filter}$ is therefore a boolean value. Consequently, whenever this variable evaluates to $true$ both sub-components of the LV system update their synaptic strengths ($\omega_t$) following an expression similar to Equation 3.2, but based on the discrepancy between actual and CS associated outcomes:

$$\Delta\omega_t = \varepsilon \times (PV_e - LV_e) \times \mathbf{u}(S_t) \tag{3.13}$$

Where $\mathbf{u}(S_t) = u_1(S_t), u_2(S_t), \ldots, u_n(S_t)$ is the real-valued vector of stimuli. Similar to the PV system, the LV system's output that then contributes to phasic dopamine activity is defined as:

$$\delta_{LV} = LV_e - LV_i \tag{3.14}$$

The learned decay mechanism, introduced by Hazy et al. (2010) for the NV system, update the synaptic strength ($\omega_t$) of its underlying neural population using the following equation:

$$\Delta\omega_{t+1} = \omega_t - \varepsilon \times NV_t \times \mathbf{u}(S_t) \tag{3.15}$$

With $NV_t$ the vector containing the current novelty value for each stimulus.

Bringing it all together the outputs from the PV, LV and NV systems are combined and contribute to a global error signal, which correspond to the phasic activity of the dopamine system, characteristic of reward expectation and delivery in classical conditioning experiments. Hazy et al. (2010) hypothesized that whenever a primary reward is expected or actually delivered, then the PV system should drive dopamine output. Otherwise, the LV and NV systems should be the main driving forces. This leads to the following equations for the global phasic dopamine output:

$$\delta_{DA} = \begin{cases} \left(\delta^{t}_{PV} - \delta^{t-1}_{PV}\right) & \text{if } PV_{filter} \\ \left(\delta^{t}_{LV} - \delta^{t-1}_{LV}\right) + \left(NV^{t} - NV^{t-1}\right) & \text{otherwise} \end{cases} \tag{3.16}$$

Where $t$ indexes the current trial and $t-1$ the previous one.

In the original paper introducing the PVLV architecture, O'Reilly et al. (2007) briefly describe the mapping between the sub-components of both PV and LV systems, and areas of the animal's brain. Hazy et al. (2010) later further developed those hypothesis using empirical results from classical conditioning and electro-physiological experiments to support their claims. What follows is a description of the mapping they established:

- $\mathrm{PV_e}$: The lateral hypothalamic area (LHA) has afferent connections from a broad variety of primary sensory areas. Moreover, a wealth of empirical evidences point to the fact that neurons of the LHA display a sustained activity in response to receiving excitatory signals (Hazy et al., 2010; O'Reilly et al., 2007). Owing to both its direct and indirect (via the penduncolopontine tegmental nucleus, PPT) projections to the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc), the LHA is in an ideal position to react and therefore signal primary reward within the animal's brain.

- $\mathrm{PV_i}$: Learning to inhibit the expected primary reward associated with a US, is the role attributed to the striosome/patch-like neurons of the ventral striatum (VS). Electro-physiological recordings show that these neural units exhibit a ramp-like pattern of firing, which peaks at the exact time of US-onset. Since they send inhibitory projections to both dorsal and ventral tiers of the dopamine system (unlike its dorsal striatum counter part), it seems reasonable to suggest that patch-like neurons are responsible for suppressing dopamine bursts at US-onset after training.

- $\mathrm{LV_e}$: The central nucleus of the amygdala (CA) receives projections from a wide variety of cortical areas. It sends both direct, as well as indirect (via the PPT) connections to the VTA and SNc. Although, it is still unclear if the CA has a direct excitatory effect or an indirect disinhibitory one over dopamine neurons, evidences suggest that sustained activity from the CA results in phasic bursts of dopamine. Furthermore, electro-physiological recordings from classical conditioning experiments have shown that CA neurons initially responsive to US alone, are able to learn to activate for a CS reliably paired with a US (as demonstrated in Chapter 4, but see Hazy et al. (2010); O'Reilly et al. (2007) as well).

- $LV_i$: Since it plays a role similar to the PVi, albeit at CS-onset, the inhibitory component of the LV system is supposed to map to a sub-population of the patch-like neurons of the VS. Empirical evidences support the idea that there exist two sub-populations of the striosome cells. The first one showing signs of activity at US-onset, whereas the second displays sustained firing at CS-onset. Therefore, it would be possible for the VS to support the inhibitory components of both PV and LV systems.

Even though, the PVLV architecture has been introduced on its own, it is in fact only half of a larger project detailing the mechanisms by which decision-making is implemented in the brain. As will become apparent at the end of the next part, the PVLV describes a system within the brain, which loosely plays a role similar to that of the '*critic*' defined earlier (Hazy et al., 2010; Maia, 2009; Sutton, 1984).

### The basal ganglia: deciding what to do next?

Similar to the current fate befalling the concept of emotion, until recently there was little consensus as to the role of the '*basal ganglia*' within the brain (Redgrave et al., 1999). However, a recurring idea born from empirical results and lesion studies, was that the basal ganglia implemented some sort of selection mechanism. This idea was expended upon by many research projects (Atallah, Frank, & O'Reilly, 2004; Berke, 2018; Frank et al., 2001; Grillner et al., 2013; Gurney et al., 2001a; Gurney, Prescott, & Redgrave, 2001b; Hazy, Frank, & O'Reilly, 2006; Houk, 2011; Maia, 2009; O'Reilly & Frank, 2006; O'Reilly & Munakata, 2013; Prescott et al., 2002; Redgrave et al., 1999; Vitay & Hamker, 2010). Leading to the now standard view of the basal ganglia complex acting as a gating mechanism, managing access to motor and cognitive resources. Here, two such projects are explored. Each adopting a hybrid approach: a top-down constraints analysis of the problem, followed by a bottom-up anatomical investigation of the brain areas part of the basal ganglia complex and their interactions. Even though, Redgrave, Prescott, and Gurney (1999) focused on the '*action selection problem*', while Frank, Loughry, and O'Reilly (2001) explored the role of the basal ganglia and pre-frontal cortex in relation to working memory, both projects devised very similar computational models and mechanistic theories.

Although there is little agreement about what working memory is, neuroscientists still agree that both the pre-frontal cortex (PFC) and basal ganglia (BG) are involved in its implementation (Atallah et al., 2004; Frank et al., 2001; Hazy et al., 2006; O'Reilly & Frank, 2006; O'Reilly & Munakata, 2013). Furthermore, Frank et al. (2001) give the following broad definition for the functionalities attributed to the working memory system:

> "*Working memory can be defined as an active system for temporarily storing and manipulating information needed for the execution of complex cognitive tasks.*"
> — Frank et al. (2001, p. 138)

From this general description, two main constraints can directly be extracted. The fact that working memory has to be an information storage, implies that it needs a mechanism for the '*robust maintenance*' of neural activity (encoding the memory), even after the initiating stimulus has disappeared. Furthermore, since information is only temporarily stored and should be available for further processing by other brain areas, the second constraint is that new incoming data should be able to quickly displace the content of working memory whenever deemed important (this is referred to as '*rapid updating*' by Frank et al. (2001)). Consequently, there is a need for a second system to decide when and which stimuli should be maintained in working memory. Since the brain is made up entirely of neurons communicating via synaptic connections, it stands to reason that information can only be stored either by modifying the synaptic weights or as a recurrent pattern of activity. Assuming that working memory is supported by a single weight-based system, both constraints would be incompatible. Indeed, a robust maintenance of the information is only possible if synapses have a slow learning rate. On the other hand, rapid updating calls for a fast learning rate, which enables the neural population to retain information presented only once. Therefore, Frank et al. (2001) suggest working memory to be implemented in an activation-based memory. Moreover, owing to its biological characteristics (slow learning rate, sparse encoding and sustained activity), it seems that the PFC is best suited to assume the role of robustly maintaining patterns of neural activity. Finally, because the BG and PFC are known to be interdependent and share the burden of implementing the functionalities attributed to the working memory system, Frank et al. (2001) hypothesized that the BG implement a gating mechanism. Thus, granting other brain systems access to working memory, while learning when to update its content.

On the contrary to working memory, the problem of '*action selection*' or '*decision-making*' has recently been explored in depth. This is in great part due to the need for controlling ever more sophisticated robots interacting with humans in the real world. In its more abstract form the action selection problem is defined by Redgrave et al. (1999) as:

> *"A "selection problem" arises whenever two or more competing systems seek simultaneous access to a restricted resource. For example, in a vertebrate or in a robot, multiple sensorimotor systems may require access to the limited resource that is the final common motor path. . . . Conflicts can also arise in domains where behavioural expression is more indirect, for instance between systems competing for access to limited cognitive resources."* — Redgrave et al. (1999, p. 1009)

Their main hypothesis is that the basal ganglia implement such a mechanism, selecting a single behavioral program at the expense of others. Furthermore, according to Redgrave et al. (1999), it is expected of a selection mechanism to execute the program with the highest support. Whenever this program ends successfully, or if it proves inefficient, the selection mechanism should switch to the next best program. Finally, if a competitor is able to garner enough support, the behavior currently executing should be interrupted and access granted to this competitor. To avoid having to deal with a wide range of command signals, sent by the different control systems to support the competing behaviors, Redgrave et al. (1999) suggest that the basal ganglia selects the winner based on a common currency. Since the behavioral programs each depend on a specific set of '*causal factors*' (of both intrinsic and/or extrinsic nature), Redgrave et al. (1999) propose that each control system computes a kind of urgency value, which they refer to as '*salience*'. Provided adequate weights those saliences can be used by the basal ganglia to select the winning program. Consequently, because the basal ganglia has no knowledge as to the nature of the control system's command, the problem of action selection can be reduced to one of '*signal selection*'. Furthermore, it also implies that the selection mechanism, implemented by the basal ganglia, is not restricted to gating access to motor resources. Indeed, similar to Frank et al. (2001) theory, the basal ganglia is hypothesized to manage both motor and cognitive resources. In addition to the constraints described so far, Redgrave et al. (1999) introduce three '*desirable*' features for a robust selection mechanism. These are:

- '*Clean switching*': This alludes to the fact that the competition between multiple control systems, to access a given resource, should be resolved rapidly and decisively in favor of the program with the most support.
- '*Absence of distortion*': Once the competition is resolved, the wining program should be executed without any interference from the losing control systems.
- '*No dithering*': It describes the behavior of a selection mechanism in the case where two competitors are closely matched. Execution of the winner will reduce its salience, until it

is below that of the next best program. Following expectations, the selection mechanism will switch to the next program, reducing its salience. Consequently, the mechanism will keep switching back and forth between the two programs, wasting energy to achieve very little.

According to Redgrave et al. (1999) a system implementing a '*winner-takes-all*' strategy fulfills the first two desirable characteristics. The last one being achievable using a non-linear positive feedback loop.

Despite the fact that Redgrave et al. (1999) and Frank et al. (2001) started in two different places, one in working memory and the other in action selection, by the end of their top-down analysis, they managed to reach similar conclusions. Thanks to Redgrave et al. (1999) we now know that, any mechanism answering the decision-making problem should implement a '*winner-takes-all*' strategy. Clearly the gating mechanism, that Frank et al. (2001) suggest the basal ganglia to be implementing, fits this requirement. The question that now remains is: how do the different brain areas, part of the basal ganglia complex, interact to support this gating behavior?

The brain areas usually considered as being part of the basal ganglia complex, and the focus of this bottom-up analysis, are:

- the striatum, which is made of the caudate nucleus, the putamen and the ventral striatum. The caudate and putamen, both sub-divisions of the dorsal striatum, are viewed as being part of the basal ganglia's input module.
- the sub-thalamic nucleus, which makes up the rest of the basal ganglia's input module.
- the globus pallidus, which can be further sub-divided into internal and external segments, usually labeled GPi and GPe respectively. The GPi being one of the major output areas of the basal ganglia.
- finally, the substantia nigra, that can also be split into two components. The first, which is referred to as '*substantia nigra pars compacta*' (SNc), is a major site of dopamine activity. While the second, called '*substantia nigra pars reticulata*' (SNr), is considered as an other output area of the basal ganglia (in addition to the GPi).

For the different control systems, each supporting either a motor or a cognitive action, it is essential to be able to send the details of the command directly to the motor or cognitive system. Indeed, since the basal ganglia only uses the common currency that is the command's salience, direct connections between control and executive systems are required. Both Frank et al. (2001) and Gurney et al. (2001a) suggest the control systems to be contained within the frontal cortex. The content of each program is then sent from the frontal cortex to the thalamus, where it is redirected toward the appropriate brain area. O'Reilly and Munakata (2013) and Frank et al. (2001) reason that any control a system wishes to exert over motor behavior has to be done via working memory. The same being true for interaction between cognitive systems. Therefore, rather than redirecting each command to a specific brain area, the thalamus connects bi-directionally with the PFC. Hence, establishing sets of thalamocortical loops, each enabling communication between a control and an executive system. To play its role as a gating mechanism, the basal ganglia receive the saliences supporting each of the behavior '*suggested*' by the frontal cortex. The projections from the frontal cortex onto the striatum, main input area of the basal ganglia, are organized into strip-like patterns (Atallah et al., 2004; Frank et al., 2001; Gurney et al., 2001a, 2001b; Hazy et al., 2006; O'Reilly & Frank, 2006; O'Reilly & Munakata, 2013). Each strip having little influence over the activity of its neighbors, it isolates the support each behavior receives from its control system. After processing this information, the basal ganglia then selects which thalamocortical loop remains active, by sending inhibitory projections onto different neural populations within the thalamus. To explain how it is possible for a single neural population (i.e.: a control system) to provide both the content of and the salience for its corresponding program, Redgrave et al. (1999) rely on the '*dual population coding*' model. In this model the detail of the program is encoded in terms of the distribution of cell activity within the population, while the salience is derived from the overall intensity of the neural activity.

Inside the basal ganglia, most of the connections in-between regions are inhibitory in nature, with the exception of the sub-thalamic nucleus which provides diffuse excitation to both segments of the globus pallidus (GPi and GPe) as well as to the SNr (Gurney et al., 2001a; O'Reilly & Munakata, 2013; Redgrave et al., 1999). Furthermore according to Gurney et al. (2001a) and Frank et al. (2001), the neural population within the striatum can be further sub-divided depending on the type of dopamine receptor as well as the projection site of each neural unit. Indeed, the first sub-population is characterized by D1-like dopamine receptors and sends inhibitory connections directly to the GPi and SNr. On the contrary, the second sub-population is equipped with D2-like dopamine receptors, while targeting principally the GPe. In turn the GPe connects via inhibitory projections directly onto neurons of the GPi and SNr. The major difference between these two families of receptors is in their interaction with dopamine. An

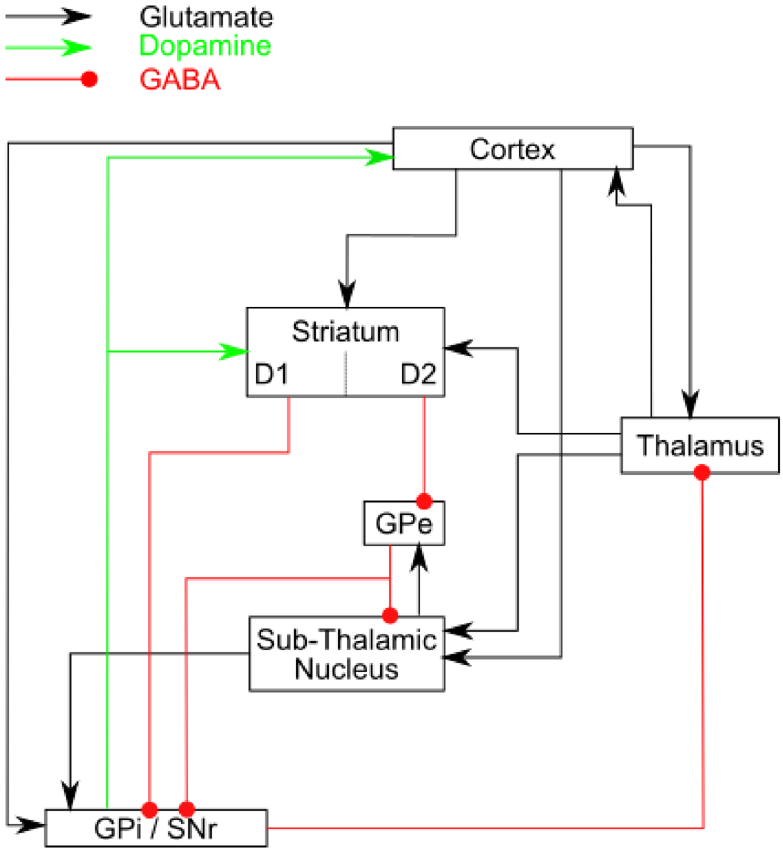


Figure 3.7 *This figure illustrates a standard representation of the gating mechanism implemented by the basal ganglia (BG). The central part of the gate relies on the fact that neurons within the dorsal striatum (DS) can be split into two sub-populations: those with D1-like dopamine receptors and those with D2-like receptors. While dopamine has an excitatory effect on units with D1-like receptors, it has an inhibitory impact on the second population. From the striatum, two pathways extend to the cortex and through the thalamus. The first pathway consist in the D1-like sub-population, sending inhibitory connections to the internal segment of the Globus Pallidus ($GP_i$). In turn, the $GP_i$ exerts an inhibitory influence over the thalamus. This pathway is usually referred to has the 'Go' pathway, since an increase in $DS1$ activity results in an increase in thalamic activity as well. The second, often called '*NoGo*', pathway originates in the D2-like sub-population. Those neurons have inhibitory projections to the external segment of the Globus Pallidus ($GP_e$). Then the $GP_e$ connects, via inhibitory kinks, to the $GP_i$. Consequently, a spike in $DS2$ activity, decreases the activation of $GP_e$ neurons. This in turn, allows neurons within the $GP_i$ to be more active and increase their tonic inhibition of thalamic activity.*

Source: Scholarpedia.org, Author: Eugene M. Izhikevich.

increase further excites neurons belonging to the D1-like sub-population, while inhibiting the activity of D2-like units. The reverse holds true: a dip in dopamine inhibits D1-like neurons and excites D2-like units. Consequently, the standard view and the one adopted by Frank et al. (2001) is to separate the basal ganglia into two processing pathways. The '*Go*' pathway consists in the D1-like sub-population, which directly inhibits GPi and SNr neurons. Hence, a phasic burst of dopamine increases the activity of D1-like neurons, inhibiting both GPi and SNr. As a result, the tonic inhibition over the selected thalamocortical loop is lifted and the corresponding program executed. The '*NoGo*' pathway is made of the D2-like striatal sub-population projecting to the GPe, via inhibitory links. A dip in dopamine levels increases the activity of D2-like neurons, therefore indirectly disinhibiting the GPi and SNr. As a consequence the targeted thalamocortical loop is further inhibited, thus preventing its program from being executed.

It should be noted that even though Frank et al. (2001) used the standard '*dual pathways*' model as a basis to explain how the basal ganglia implement a gating mechanism, Gurney et al. (2001a) suggested a slightly different interpretation. In their point of view, the pathway consisting in the D1-like sub-population connected to the GPi and SNr, is responsible for the actual '*selection*'. On the contrary, the D2-like and GPe pathway, having no influence outside the basal ganglia, assumes a role akin to a '*control*' system. Still the overall mechanism implemented by the basal ganglia complex remains similar and both explanations fit within the requirements extracted during the top-down analysis.

In conclusion, part of the learning process happens through conditioning. In other words, it means that a previously insignificant stimulus can acquire reward predicting capabilities by association with a stimulus natively signaling primary reward. According to O'Reilly et al. (2007) and Hazy et al. (2010), this mechanism is implemented by three systems within the ventral striatum, amygdala, lateral hypothalamus and both ventral tegmental area and subtantia nigra pars compacta, part of the dopamine system. The dopamine activity being the main output signal, it is used to influence learning inside and outside the architecture, and to critique any decisions made. Although the role of the basal ganglia was not established a decade and a half ago, it seems that projects and architectures, like that of Frank et al. (2001) and Redgrave et al. (1999), have helped cement its role as a decision-making system. Using the dopamine signal generated by the critic, the basal ganglia, especially the dorsal striatum, choose how to act by influencing the different thalamo-cortical loops connecting control systems to executive ones. Hence, the gating mechanism implemented within the basal ganglia's dual-pathways, can be compared to a higher-level control system. This idea of the brain being a hierarchy of control systems and distributed memory is beyond the scope of this project. Furthermore, it has already been discussed at length by Hawkins and Blakeslee (2007). Finally, as Maia (2009) suggested, given the brain structure described so far it possible to conceive of the actor and critic as mapping directly on the striatum. The dorsal part being a major component of the actor, while the ventral section is at the core of the critic. However, it should be noted that a number of attempts have been made to identify the brain areas or neural sub-populations of the striatum involved in reinforcement learning and the actor-critic model ( Joel, Niv, and Ruppin (2002) make a selective review of several such models, but see O'Reilly and Frank (2006), Gurney et al. (2001a), Suri and Schultz (1999), Niv (2009), O'Reilly et al. (2007), O'Doherty, Lee, and McNamee (2015), Collins and Frank (2014), Hazy et al. (2010), as well as O'Reilly and Munakata (2013) for other examples).

# Summary

*The most basic form of learning within the brain is classical conditioning. Classical conditioning describes a process by which a conditioned stimulus (with no particular value for the individual) can acquire significance by association with an innately relevant stimulus. Going one step further, through instrumental conditioning the likelihood of a rewarding action to be performed again, should a similar situation arise, is increased. On the contrary, an action which leads to a punishment is less likely to be executed another time. The fundamental principles guiding the instrumental conditioning process have been mathematically modeled by R. A. Rescorla and Wagner (1972), and later extended into a learning paradigm better known as '*reinforcement learning*' by Sutton and Barto (1998). In the animal's brain, the decision-making system, which includes learning, is split into two sub-systems: a critic and an actor. The critic part is responsible for predicting the error in reward or punishment. According to O'Reilly, Frank, Hazy, and Watz (2007) and Hazy et al. (2010), it mainly consists of two mechanisms the '*Primary Value*', which learns to expect the occurrence of an unconditioned stimulus, and the '*Learned Value*', which learns to predict the reward associated with conditioned stimuli. Both mechanisms are connected to the ventral tegmental area, and, thus, directly drive the activity of the dopaminergic system. The actor, on the other hand, based on the signals received from the critic adapts its policy. Furthermore, it relies on the gating mechanisms implemented within the basal ganglia to select which action to perform next given the current situation. The dorsal striatum can be split in two populations of neurons, depending on their affinity to dopamine. Indeed, within the dorsal striatum some neurons are excited by an influx of dopamine (they are referred to as D2-like units), while the others are inhibited by dopamine (they are called D1-like neurons). The D1-like population connects directly to the internal segment of the globus pallidus (GPi) and from there to the thalamus. A burst of dopamine activity will inhibit the D1-like neurons, hence, releasing the tonic inhibition the GPi has over the thalamus. D2-like neurons, on the other hand, project to the external segment of the globus pallidus (GPe), which sends inhibitory connections to the GPi. Therefore, an increase in dopamine will result in an increase of activity in the D2-like population, and, thus, a further inhibition of the thalamus, and vice versa for a decrease in dopamine concentration. The thalamus being bi-directionally connected to the pre-frontal cortex, its activation or inhibition will allow or deny, respectively, the command from the pre-frontal cortex to be forwarded to the motor system. The brain mechanisms responsible for the entire decision-making process have been described by both Redgrave et al. (1999) and Frank et al. (2001).*

# CHAPTER 4

# Recreating: An anatomically constrained neural network model of fear conditioning

When working on their research project Armony et al. (1995) had two objectives in sight. As with any other scientific undertaking the main goal was to validate a set of '*principles*', as they are referred to by Armony et al. (1995) (the reasons for using the term '*principles*' rather than hypotheses or research questions will become clear in the coming section), and explain two independent sets of findings. However, hidden behind this first standard aim was the desire to highlight the usefulness of computational models in helping neuroscience move forward. Certainly, this is a cause this thesis fully adheres to. By their very nature computational models are the perfect tool to validate the mechanisms that brain systems are suggested to implement. Furthermore, once the validity of a computational model has been established it can serve as an abstract stand-in for the animal's brain. Hence, allowing researchers to easily perform lesion studies and make predictions that can later be more efficiently validated by neuroscience (a fact exemplified by Armony, Servan-Schreiber, Cohen, and LeDoux (1997); Armony, Servan-Schreiber, Romanski, Cohen, and LeDoux (1997)). Effectively, establishing a feedback loop between computational models and neuroscience.

Another impetus for re-creating the experiment and re-implementing the neural network described by Armony et al. (1995), is that the model of the '*two-pathways*' to the amygdala suggested by LeDoux (1986, 1990, 1992, 1998), and Romanski and LeDoux (1992) has been integrated within the **PrimEmo** architecture. Furthermore, even if only half of this model appears in the **ProtoEmo** architecture it can be used to make predictions as to the capabilities of the amygdala to accurately distinguish between similar stimuli, such as two tones with adjacent frequencies for example.

The sections making up this chapter have been organized to mirror the layout of the original paper from Armony et al. (1995). As such, the first section introduces the model of the two pathways to the amygdala suggested by LeDoux (1986), before describing the neural structure implemented by Armony et al. (1995) to validate said model. The next section details the fear conditioning experiment used as a testing ground. The third section presents the results gathered by re-implementing both the neural network and the conditioning experiment. Finally, the last section discusses the differences between the results published by Armony et al. (1995) and those collected in this re-creation.

## 4.1 A neural network constrained by fear

As stated above, the main goal of Armony et al. (1995) was to test the ability of three principles to explain two sets of behavioral and physiological findings. To do so, Armony et al. (1995) designed and implemented a neural structure based on the following assumptions:

1. Processing units: populations of real cells coding for the same piece of information can be abstracted as non-linear summing devices;
2. Dual pathway connectivity: sensory information describing the state of the environment is encoded by two parallel systems, before converging on the amygdala;
3. Learning: the synaptic strength for the connections between neurons is updated using a modified Hebbian learning rule, also known as the Stent-Hebb rule (Stent, 1973).

Simulating a fear conditioning experiment, Armony et al. (1995) explored whether the neural network would express conditioned responses similar to those observed in animals and known to be the result of amygdala activation. At the physiological level, Armony et al. (1995) examined if the neural units displayed any change in activities as a result of conditioning analogous to those measured in single-cell recordings.

### 4.1.1 Two pathways to the amygdala

In an effort to fully describe the neural circuitry underlying the emotion of fear, as well as the mechanisms by which animals learn what to be afraid of, LeDoux (1986, 1990, 1992, 1998) took advantage of the classical conditioning paradigm. Fear conditioning is a favored method of investigation for several reasons. For one it is easily repeatable, since fear can be induced in most animals via a simple foot shock. Further, in the post-conditioning test phase, the stimulus triggering an adaptive reaction is known and, therefore, its point of origin in the brain is more easily identifiable. Indeed, by definition (see Chapter 3 for more details) conditioning is a process through which an animal learns to associate a CS (like a tone) with the rewarding or punishing value of a paired US (a foot shock for example). Both the CS and US are chosen and, thus, known by the experimenter. Consequently, to trace the paths through which the CS is transmitted within the brain, one can start by investigating the brain systems in charge of processing information of similar modality.

Considering that all of his experiments involved the conditioning of animals to a sound, LeDoux let himself be guided by the natural flow of information (LeDoux, 1986, 1990, 1992, 1998; Romanski & LeDoux, 1992). Thus, setting the starting point of his investigation of the neural circuitry of fear in the auditory cortex. Through a first lesion study, complete removal of the auditory cortex proved to have no significant impact on the conditioning of rats (LeDoux, 1986, 1990, 1992, 1998; Romanski & LeDoux, 1992). Therefore, following the flow of information LeDoux decided to conduct a second study targeting the auditory portion of the thalamus. Removal of the '*medial geniculate body*' (MGB), an area of the thalamus responsible for the transmission of data related to sound, proved to be detrimental to conditioning. In fact, after removing the MGB, animals were unable to learn to be afraid of a tone, even when paired with a foot shock. Surprisingly, however, they were still capable of reacting to other innately relevant stimuli, such as the foot shock itself. Using chemical tracing methods, it was determined that the auditory thalamus apart from sending connections to the cortex projected significantly to four other sub-cortical areas too. Severing the link between the thalamus and the first three regions yielded no results concerning conditioning. However, cutting the path from the thalamus to the amygdala, resulted in the same interferences in the conditioning process, as those observed after lesioning the thalamus.

Further conditioning experiments and lesion studies, involving not just a single, but two similar tones, allowed (LeDoux, 1986, 1990, 1992, 1998) to conclude that: *1*) there are two pathways from the thalamus to the amygdala, one directly connects the two areas together, whereas the other involves sensory related cortical areas; *2*) if a simple stimulus, a tone for example, is used as a CS, then both pathways are equivalent; *3*) however, in more complex cases, such as conditioning using two similar tones, one of them paired with a US, the thalamo-amygdala pathway is incapable of differentiating between the two tones and reacts similarly to both. Therefore, the indirect pathway through the cortex while slower provides more detailed information concerning the CS. Thus, allowing animals to learn to react to one tone, but not the other.

### 4.1.2 Model

The model suggested by Armony et al. (1995); Armony, Servan-Schreiber, Cohen, and LeDoux (1997); Armony, Servan-Schreiber, Romanski, et al. (1997) is split into two modules, each corresponding to one of the pathways to the amygdala. As shown in Figure 4.1, the first module includes the medial section of the MGB (MGm) and a layer representing the amygdala. While the second module is made of the ventral section of the MGB (MGv) and the cortex, which receives data from both MGm and MGv, and then projects to the amygdala. Every layer is populated with non-linear computational units, whose activation value can be thought of as the average neural activity of a small population of real neurons coding for the same patterns. The processing units used for this model strike a balance between capturing the basic features of the response of a real cell and being computationally tractable. For each unit, its output remains null as long as its net input is below the threshold $x_{thr} = 0$, and it saturates upon reaching the cell's maximum firing rate, $x_{sat} = 1$:

$$act(x) = \begin{cases} 0 & \text{if } x - x_{thr} < 0 \\ x - x_{thr} & \text{if } 0 < x - x_{thr} < x_{sat} \\ x_{sat} & \text{if } x_{sat} < x - x_{thr} \end{cases} \tag{4.1}$$

Where, in Figure 4.1, one layer projects to another, the author chose to fully connect every sending units to each of the receiving units. Therefore, the net input of any neuron is computed as the weighted sum of its sending units' activation values:

$$net_r = \sum_s a_s \times \omega_{rs} \tag{4.2}$$

Where $\omega_{rs}$ is the synaptic weight between the sending unit $s$ and the current receiving unit $r$.

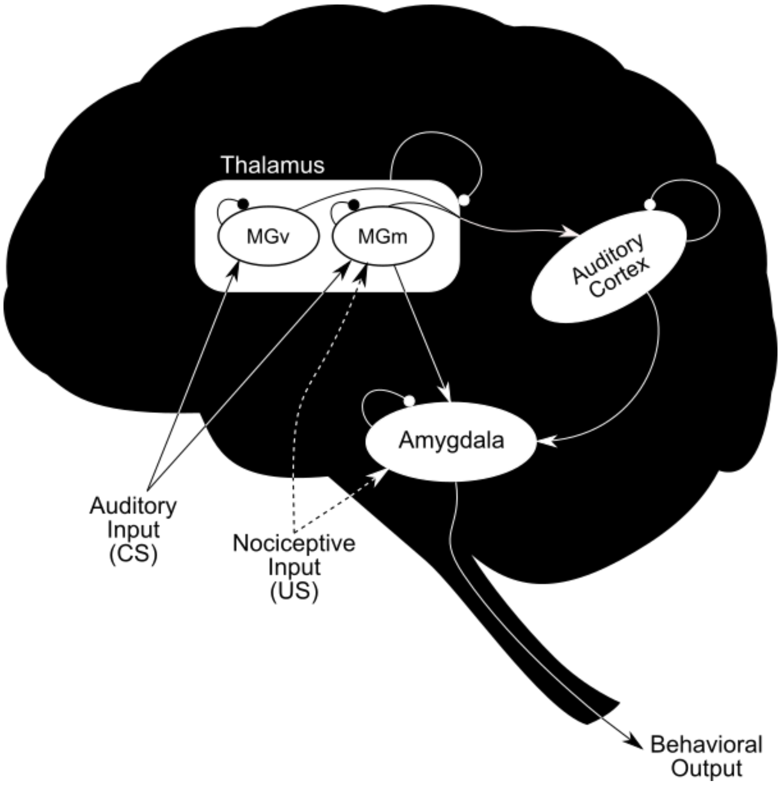


Figure 4.1 *Using the architecture presented above, Armony et al. (1995) sought to validate the model of the* ‘two-pathways to the amygdala’ *suggested earlier by LeDoux (1986, 1990, 1992), as well as Romanski and LeDoux (1992). According to this model, sensory information entering the thalamus and destined for the amygdala takes two paths. The first one, links directly to the amygdala and provides fast communication, but at the cost of precision. Indeed, the information is very general and an animal could not tell the difference between two tones with adjacent frequencies by relying on this data alone. The second path, goes through the cortex (the auditory cortex in this case) and further refines the signal. Hence, providing much more detailed information to the amygdala. In the case of a conditioning experiment, if an animal has to simply react to a tone, then both pathways are equivalent save for the time it takes for the signal to reach the amygdala (LeDoux, 1992; Romanski & LeDoux, 1992). However, if the tone the animal has to react to is very specific, then the second pathway becomes very important, since it allows the amygdala to discriminate between different frequencies. Therefore, the overall role of the first pathway is to ready the body for a* ‘fight-or-flight’ *response, while the second one helps decide if a reaction is actually necessary.*

To account for the presence of inter-neurons in real brain areas lateral inhibition has been implemented through a soft version of the winner-takes-all mechanism. Following this algorithm in each layer the winning unit (defined by strength of activation) inhibits all other units by an amount proportional to its net input:

$$\begin{aligned} a_{win} &= activ(net_{win}) \quad \text{for the winner} \\ a_r &= activ(net_r - \mu \times a_{win}) \quad \text{for the other units} \end{aligned} \tag{4.3}$$

$\mu$ was set to 0.2 for all units, in all layers, and left unchanged for the whole duration of the simulation.

In the competitive-learning scheme used for this model the number of input patterns for which a given unit is activated is inversely proportional to the number of units in the layer. Therefore, to capture the broad tuning capabilities of the first module, the MGm and amygdala layers have very few neurons (for the current implementation the number of units in the MGm and amygdala layers have both been set to $n = 3$). On the contrary, the MGv and cortical layers, both part of the second module, are populated by $n = 8$ units per layer, more than double the number of neurons. Hence, allowing units to acquire a narrower tuning which accounts for the improved categorization capabilities of the cortex.

After the presentation of each input pattern, learning is achieved through the modification of the synaptic weights of all excitatory connections, but not the inhibitory ones. For the modification of the weights, Armony et al. (1995) used a variant of the Hebbian learning rule. It is well known that the direct application of Hebb's learning rule quickly leads to saturation. Therefore, to work around this issue, the so called Stent-Hebb algorithm was applied (Stent, 1973). It allows for both increases and decreases in synaptic strength based on the correlation in activity between the sending and receiving units. The equation used for updating the weights is:

$$\omega'_{rs} = \begin{cases} \omega_{rs} + \varepsilon \times a_r \times a_s & \text{if } a_s > a_{avg} \\ \omega_{rs} & \text{otherwise} \end{cases} \tag{4.4}$$

Where $a_{avg}$ is the average activation value sampled over the units of the sending layer and $\varepsilon = 0.1$ is the learning rate. Furthermore, to account for the decrease in synaptic strength for uncorrelated sending units, for each unit in the receiving layer the sum of its input weights is kept constant through multiplicative normalization. That is, each weight is further processed using the formula:

$$\omega_{rs} = \frac{\omega'_{rs}}{\sum_s \omega'_{rs}} \tag{4.5}$$

As a consequence, the synaptic strength of sending units whose activation value remain below the layer's average will decrease, while the weights of correlated units increases.

In the simulated conditioning experiment, sound was used as the CS, while the nociceptive US was represented by a single binary unit connected to all units within the MGm and amygdala layers (as depicted in Figure 4.1). The strength of the synaptic connections between the US unit and the receiving neurons of the first module were set to 0.4 and remained unchanged for the duration of the simulation. According to Armony et al. (1995) this design decision was intended to "*capture the effect of diffuse somatosensory information associated with a US such as a footshock*". A set of training patterns was created by dividing the auditory spectrum into 16 pure tones of contiguous frequencies in an arbitrary scale. Consequently, the tones as well as the CS were represented in the simulation by overlapping patterns of activation in the input layer (as shown in Figure 4.2). In addition to the competitive learning algorithm (described in Equation 4.4 and Equation 4.5), the use of overlapping nonorthogonal patterns facilitates the development of topographical representations of the input in the model's layers.

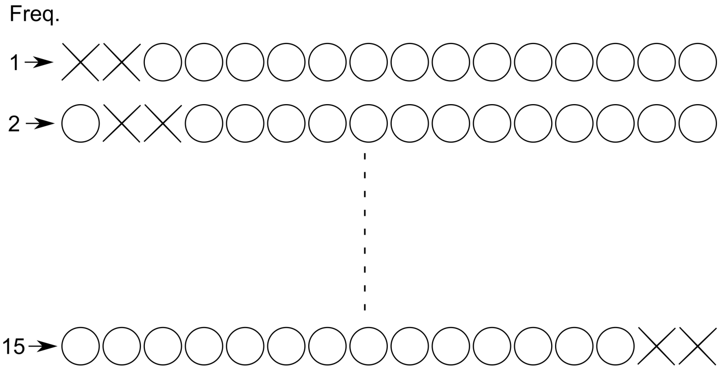


Figure 4.2 *This figure depicts the input features used in Armony et al.'s (1995) conditioning experiment. Frequencies are represented by nonorthogonal overlapping patterns of activities in the input layer. When associated with the competitive learning mechanism (described in Equation 4.4 and Equation 4.5), this type of pattern facilitates the development of topographical representations in the model's layers.*

## 4.2 Method

As alluded to at the beginning of Section 4.1, to ascertain the validity of the model of the dual pathways to the amygdala suggested by LeDoux (1986, 1992), Armony et al. (1995) simulated a fear conditioning experiment. The goal was to examine whether the neural network, whose implementation was guided by the principles expressed in Section 4.1, could reproduce the two sets of behavioral and physiological findings. At the behavioral level, it is expected that the network's behavioral output, defined as the sum total activity of the amygdala layer, should drastically increase for the frequency corresponding to the CS, while showing a net decrease in activity for all other frequencies. At the physiological level, the neural units within layers receiving information about the US, either directly or indirectly, are anticipated to have their receptive fields (RF) shift towards the CS's frequency. A neuron's receptive field is defined in this context as the frequency range for which its output activity is non-zero. The effects of conditioning should also be stronger the closer a unit's best frequency (BF; the frequency for which a neuron is maximally activated) was to the CS in the pre-conditioning phase. The fear conditioning experiment has been split into three distinct phases.

Rather than programmatically constraining the initial synaptic weights so that input patterns would be represented topographically in all layers, Armony et al. (1995) decided to randomly initialize the synaptic strengths with values in the range $[0,1]$, then let the learning algorithm do the synaptic pruning. According to Armony et al. (1995), this is to ensure that any modification in the receptive fields of the neural units is the result of conditioning alone, and not an unforeseen consequence of setting a specific set of weights. Consequently, the first, '*development*', phase consists in presenting sequentially all input patterns to the network without activating the nociceptive US unit. After each pattern, the weights of the excitatory connections were updated using the Stent-Hebb rule described in Equation 4.4, and normalized unit-wise using Equation 4.5. The loop was repeated until all units within the neural network had developed a stable RF. The curves labeled '*Pre*' in Figures 4.3 — 4.6, show standard receptive fields for units belonging to each layers.

Once all units in the neural network had developed stable topographic representations for all input patterns, the conditioning paradigm was simulated. The '*conditioning*' phase is very similar to the previous development phase. The only difference is that, a frequency was arbitrarily chosen to be the CS, and, therefore, associated with the activation of the nociceptive US unit. All input patterns were again sequentially presented to the input layers. The weights of the excitatory connections were adjusted following the same extended Stent-Hebb rule (defined in Equation 4.4 and Equation 4.5). Again, the process repeated until the RFs of all units stabilized. The curves labeled '*Post*' in Figures 4.3 — 4.6 depict the new receptive fields units developed as a result of conditioning.

Finally, the '*testing*' phase spreads across both of the previous phases. By measuring the sum total response of the amygdala's units (the behavioral response), Armony et al. (1995) investigated the overall behavior of the network in response to the CS before and after conditioning (see Figure 4.7). Additionally, for all neural units in the network their activation values for each of the frequencies were also recorded after the first two phases to analyze any change in receptive fields between pre- and post-conditioning (shown in Figures 4.3 — 4.6).

## 4.3 Results

The results presented in this section have been gathered by reproducing as closely as possible the fear conditioning experiment designed by Armony et al. (1995). The next section discusses the adequacy of my results compared to the ones published in the original paper (Armony et al., 1995).

At the beginning of the development phase, when the synaptic strengths were randomly initialized neurons responded, on average, equally but weakly to all input patterns. By repeatedly presenting the 15 tones to the network's inputs and adapting the weights of the excitatory connections using the competitive algorithm defined in Equation 4.4 and Equation 4.5, all neural units developed a receptive field. That is the output activity of a unit was non-zero for a range of adjacent frequencies. This receptive field is centered around a best frequency, which correspond to the tone for which the neuron's output is maximal. This is made clear by the pre-conditioning graphs in Figures 4.3 — 4.6. Furthermore, the receptive fields of units belonging to the amygdala and MGm layers are broad, supporting the idea that the direct pathway to the amygdala can only encode coarse-grained data. On the contrary, units from both the cortical and MGv layers have narrower receptive fields allowing for a finer-grained encoding of the information along the indirect pathway. According to Armony et al. (1995) this discrepancy in the breadth of the receptive fields is to be expected, since when using lateral inhibition the representational capability of neurons in a layer is inversely proportional to the number of units in said layer.

After conditioning, a number of units displayed significant frequency-specific changes in their receptive fields. This, however, only occurred in layers that received information about the nociceptive input. Consequently, only units in the MGv layer saw no change in their activation pattern after conditioning. The cortical layer, although not a direct target of the US binary unit, does indirectly receive data concerning the US via its connections with the MGm layer. Hence, forcing its units to adapt to the additional signal. Inside the concerned layers, units whose best frequency was close to the CS before conditioning saw a drastic increase in activity for the CS, resulting in a shift of their whole receptive field toward the CS. As a matter of fact, any unit whose activity was non-zero for the CS exhibited an increase in activity for the CS and a decrease for all other frequencies. The effects are more pronounced the closer the unit's best frequency was to the CS prior to conditioning. Which means that for units, that did not include the CS in their receptive field at the end of the development phase, no substantial change occurred. This is evidenced by the graphs showing the discrepancy in neural activity pre- and post-conditioning in Figures 4.3 — 4.6.

Finally the behavioral response, which is defined as the sum total of the amygdala's activities, produced a twofold increase in the network's response to the CS. Such a drastic change between pre- and post-conditioning output can be explained by the results presented in the previous paragraph. Indeed, given that units whose best frequency was close the CS prior to conditioning displayed a significant increase in their activity, it follows that the weighted sum of the activation values from those neurons would increase too. This aggregation of output activities also explains why the farther a frequency is from the CS the less intense the behavioral response is. As shown in Figure 4.7, this results in a curve climbing as the tone comes closer to the CS, then reaching a peak for the CS, and falling back afterwards.

## 4.4 Conclusion

The fear conditioning simulation that I re-created based on the description from Armony et al. (1995) seems to have yielded results similar to those published in the original paper. Due to the random initialization of the synaptic strengths in the development phase the match is not exact. However, the features necessary for drawing conclusions as to the validity of their principles are present in my observations as well.

In conclusion, the neural network was able to reproduce the overall behavior of the amygdala observed in fear conditioning experiments performed on animals. At the physiological level,

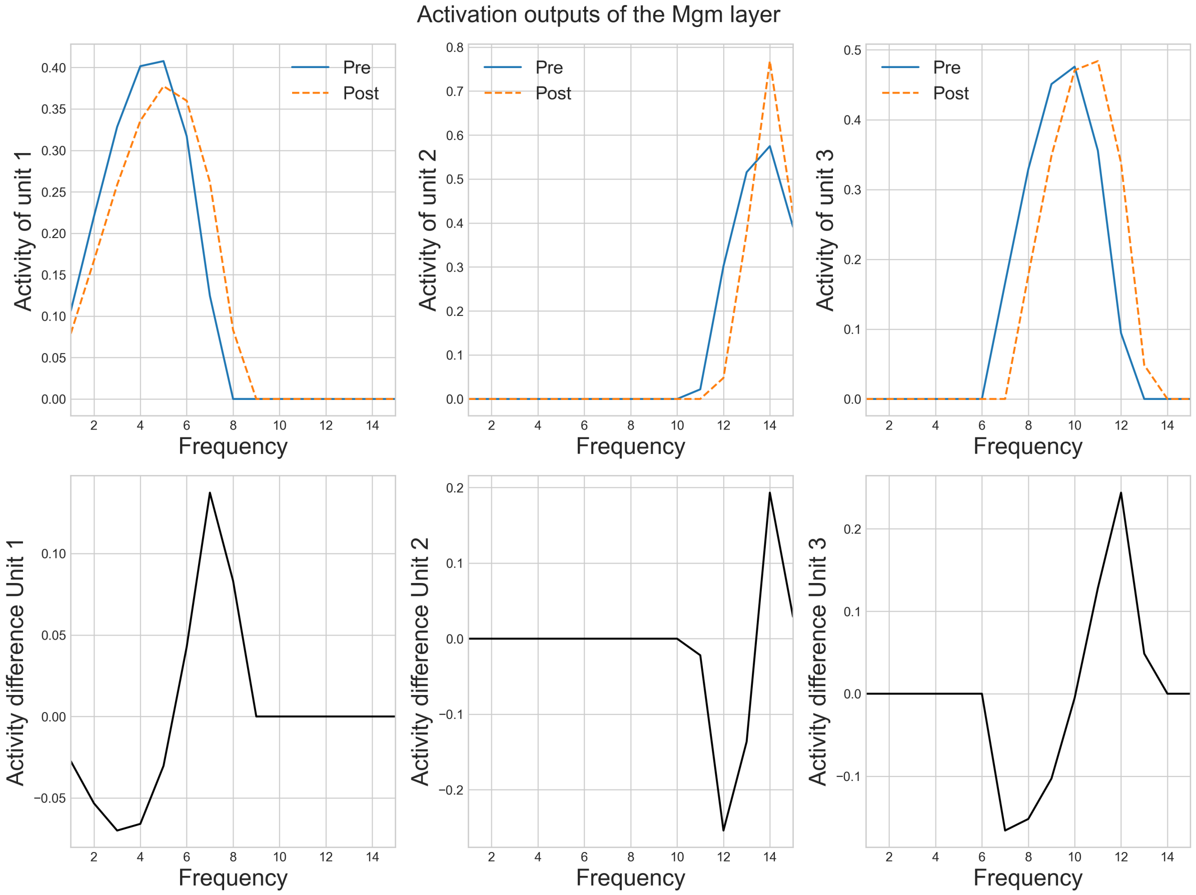


Figure 4.3 *This figure shows the activation of the neural units in the MGm layer at the end of the development phase (line labeled* 'Pre'*) and after conditioning (labeled* 'Post'*). The difference in receptive fields are quite apparent, as is the increase in activity for the unit with a best frequency close to the frequency chosen as the CS (frequency* 14 *in this case). The actual discrepancy in the activation of each unit before and after conditioning is made apparent by the set of graphs on the bottom line. As Armony et al. (1995) remarked in the original paper, the receptive field of all units shift toward the CS, this is not a phenomenon limited to only a single unit.*

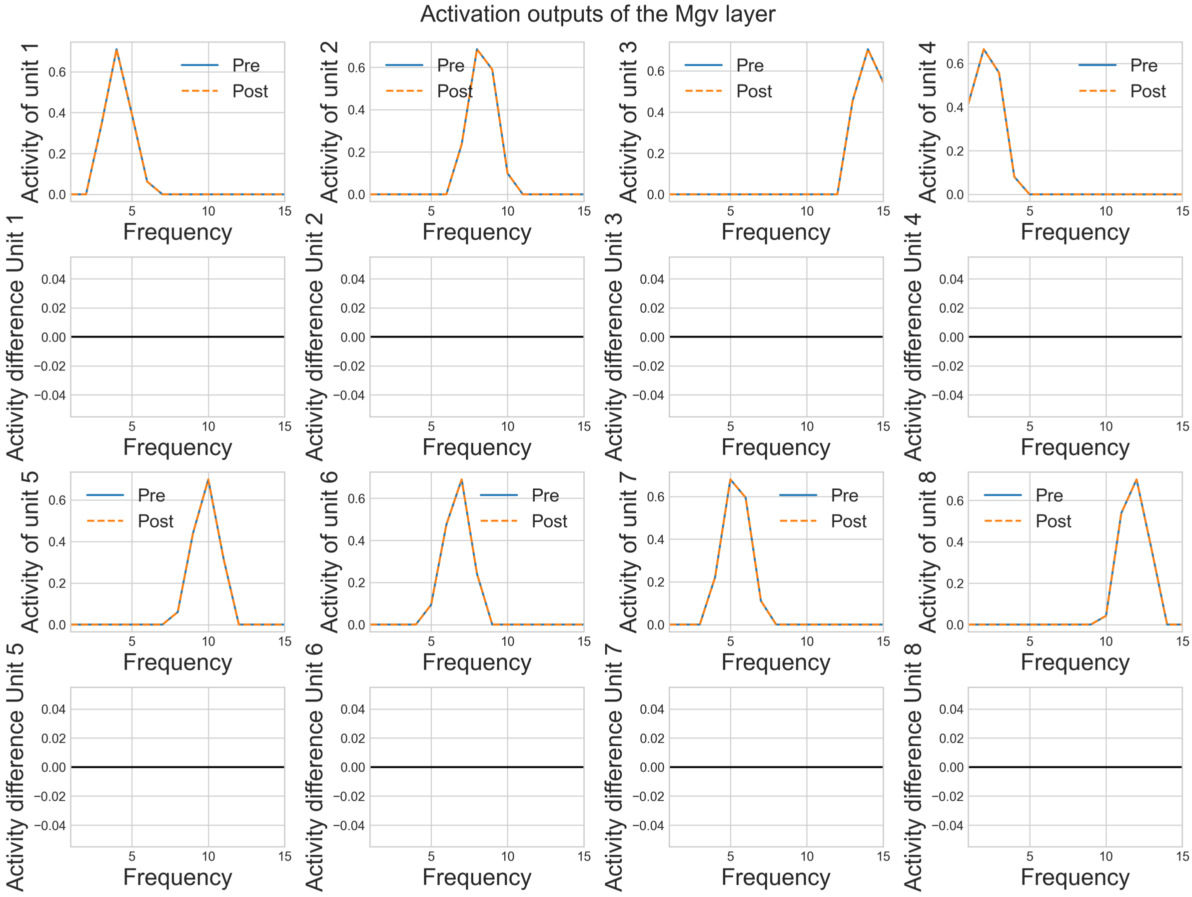


Figure 4.4 *This figure shows the activation of the neurons in the MGv layer pre- and post-conditioning. As mentioned by Armony et al. (1995), since the MGv does not receive any information about the nociceptive US the receptive fields of its neural units are not modified by the conditioning paradigm. This is made abundantly clear by the flat lines on the second and last row of this figure, which represent the difference in activity before and after conditioning for each unit.*

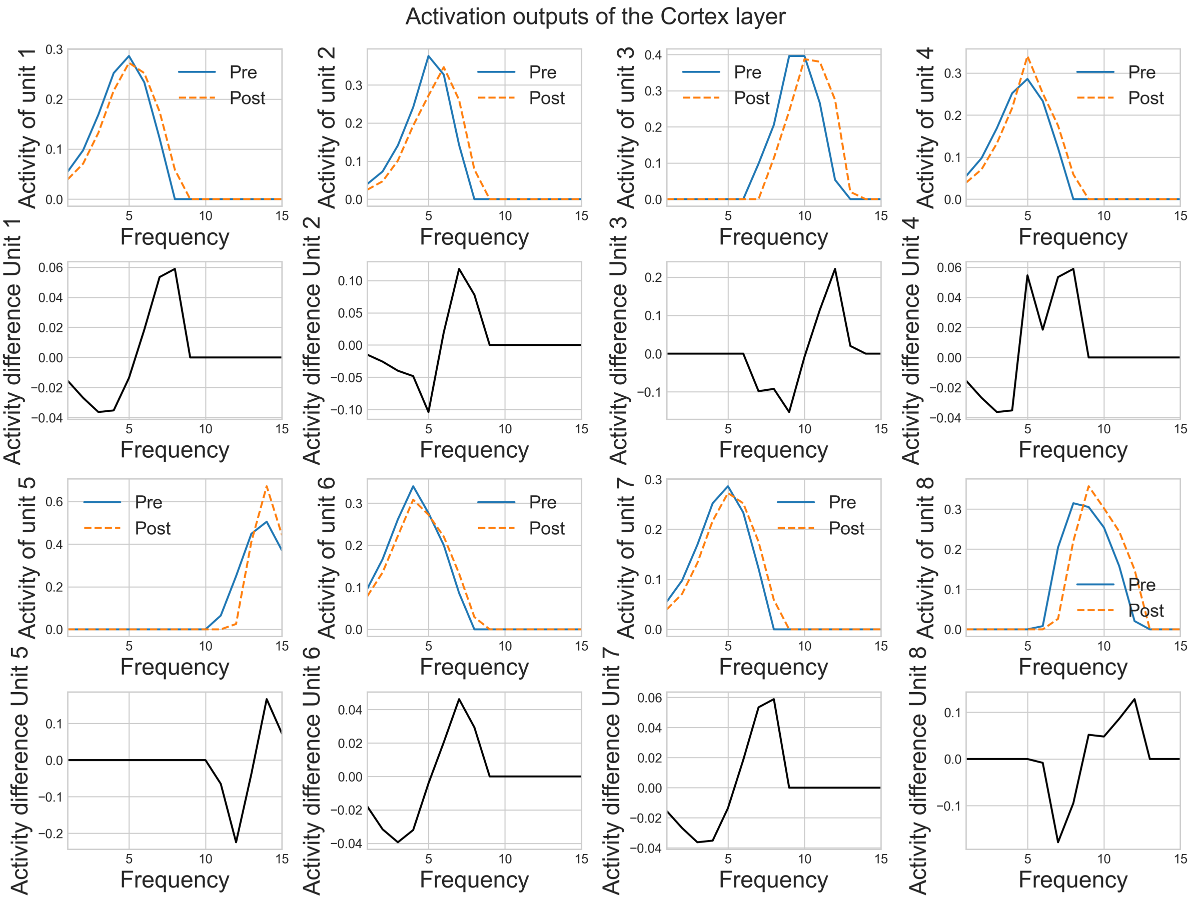


Figure 4.5 *This figure displays the activation of the neural units in the cortical layer before and after conditioning. Although, the auditory cortex does not receive any direct information about the nociceptive US, it does receive indirect data via the connections between the MGm and cortical layers. Hence, neurons inside the cortex also display a modification in their receptive field.*

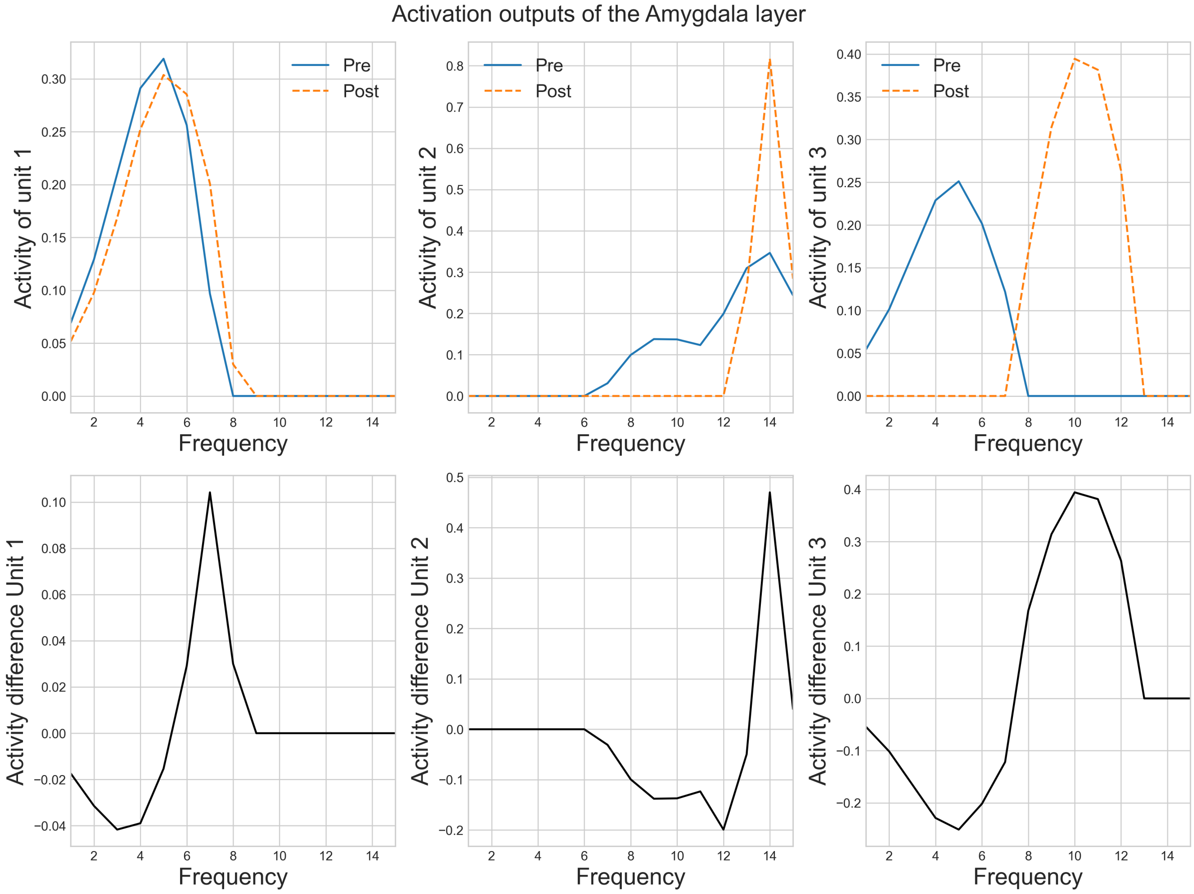


Figure 4.6 *This figure depicts the activation of the neurons in the amygdala layer. According to LeDoux (1986, 1992) the amygdala is the target of two pathways. The first connecting directly the thalamus with the amygdala provides coarse information about incoming stimuli. The second pathway, going from the thalamus to the amygdala via the auditory cortex while slower than the first pathway, sends more fine-grained stimuli related data. Hence, the thalamo-cortico-amygdaloid path allows animals to be conditioned with tone of more specific frequencies, without reacting to tones with adjacent frequencies. In the model designed by Armony et al. (1995) the amygdala, therefore, receives information directly from the nociceptive US, but also from the MGm and the cortical layers. As a result, the receptive fields of the different neural units are substantially modified after conditioning occurs.*

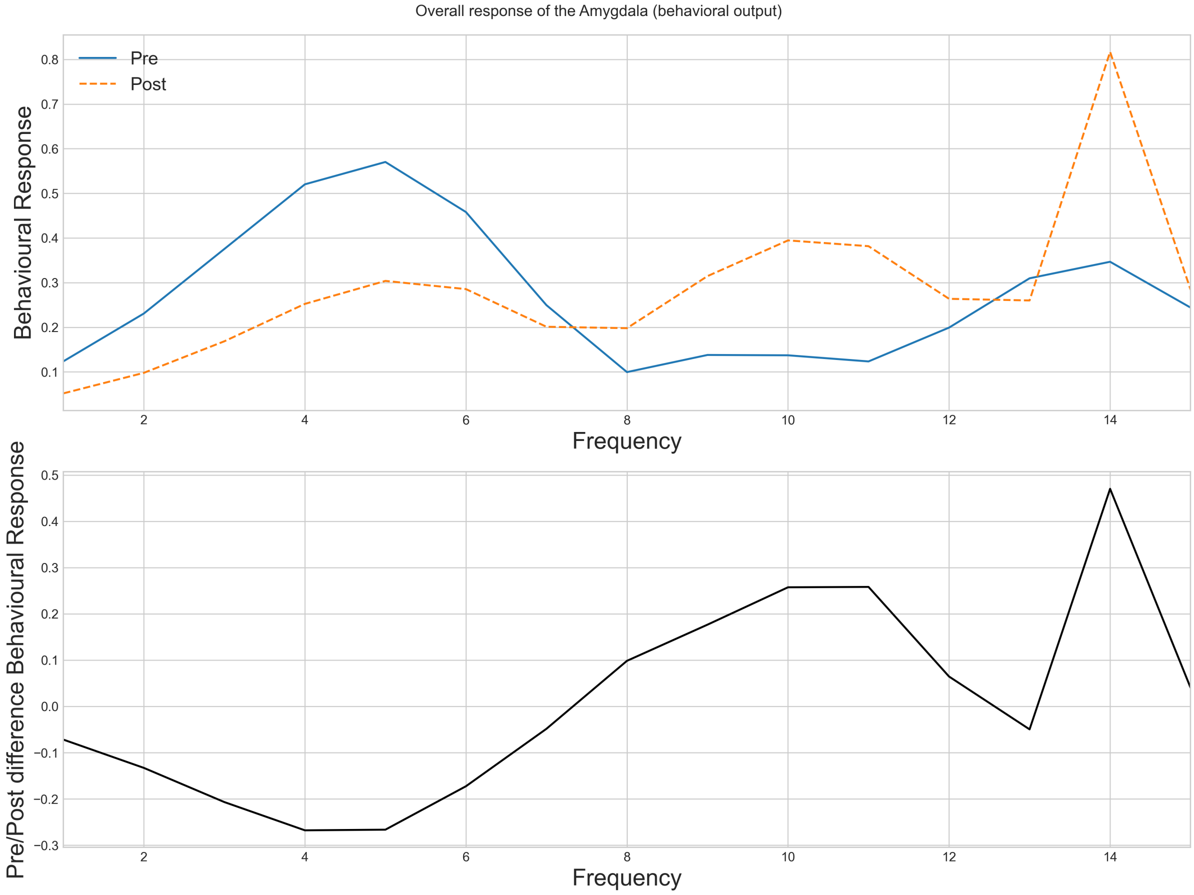


Figure 4.7 *This figure shows the evolution of the behavioral response against the different frequencies before and after conditioning. Since the behavioral response is defined as the total activity of the amygdala layer, and neurons in the amygdala see a drastic increase in activity around the CS frequency after conditioning, it follows that the behavioral response is at its maximum for the CS frequency. Activities for frequencies around the CS also see an increase in activity. However, the further a frequency is from the CS, the less it is impacted and can even decrease as a result of the shift in the receptive fields of the amygdala's neurons.*

the neural units did develop receptive fields that were subsequently altered via conditioning in a way similar to what was observed in single-cell recordings in animals (see Armony et al. (1995) for graphs depicting those empirical results). As a consequence, Armony et al. (1995) concluded that the principles that guided the design of the neural network were indeed correct. Meaning that the model of the dual pathways to the amygdala suggested by LeDoux (1986, 1992), and Romanski and LeDoux (1992) can also be considered valid. More importantly though, Armony et al. (1995) were able to show that:

> "*... a neuroanatomically constrained network, together with a biologically plausible learning algorithm, can capture important aspects of the behavioral and physiological consequences of fear conditioning. The model offers a complementary approach to experimental studies for examining issues pertaining to the acquisition and expression of fear learning in the brain.*" — Armony et al. (1995, p. 252)

Furthermore, results pertaining to the simulation of a lesion study (Armony, Servan-Schreiber, Cohen, & LeDoux, 1997; Armony, Servan-Schreiber, Romanski, et al., 1997), proved the model capable of predicting the effects of severing the connections between different parts of the network. Thus, adding their voices to the mounting evidence that computational models should be an essential tool in the belt of any neuroscientist.

# Summary

*The study presented by Armony et al. (1995) and reproduced in this chapter has two purposes. The first is to validate the three following principles:*

1. *Processing units: populations of real cells coding for the same piece of information can be abstracted as non-linear summing devices;*
2. *Dual pathway connectivity: sensory information describing the state of the environment is encoded by two parallel systems, before converging on the amygdala. This is a model suggested by LeDoux (1986, 1992), as well as Romanski and LeDoux (1992);*
3. *Learning: the synaptic strength for the connections between neurons is updated using a modified Hebbian learning rule also known as the Stent-Hebb rule (Stent, 1973).*

*Secondly, it attempts to show that computational models could be a worthwhile tool to validate mechanisms suggested by neuroscience and then inform future endeavors by making predictions. To test the validity of their three principles, Armony et al. (1995) built an anatomically constrained neural network, and tested it using a simulated fear conditioning experiment. Through this simulation they evaluated whether the network would exhibit behavioral responses characteristic of conditioned animals. At the physiological level, they examined whether the neural units making up the network would develop* 'receptive fields' *similar to those observed in single-cell recordings. Since sound was used as the CS in this experiment, the term receptive field is defined as the frequency range for which a neural unit produces a non-zero output. The outcomes of my re-creation of this experiment match very closely the results published by Armony et al. (1995). As both sets of findings were reproduced by the neural network. It follows that the three principles guiding the model's design are valid. More importantly though, this project constitutes further evidence that computational models* "offer a complementary approach to experimental studies" *Armony et al. (1995, p. 252). Finally, the model of the two-pathways to the amygdala has been integrated in the design of the* ***PrimEmo*** *architecture (see Chapter 6 as well as Appendix B). Furthermore, even though* ***ProtoEmo*** *lacks any cortical area (see Chapter 5 and Appendix A for a description of* ***ProtoEmo****'s design), the model suggested by LeDoux (1986, 1992) and validated by this study helps predict the encoding and, therefore, detecting capabilities of the amygdala.*

# CHAPTER 5

# **ProtoEmo**: Warning the brain about survival-relevant stimuli

The idea that emotions evolved to improve the chances of survival of our animal ancestors, is not novel. In fact, it has been tackled, in one form or another, by other well-known scientists such as Cannon (1929), MacLean (1949, 1952), Panksepp (1998, 2005); Panksepp et al. (1982) and Damasio (1997, 2008). It is an idea that originated from Darwin's '*The Expression of Emotions in Man and Animals*' (see Chapter 2 for a more extensive recount of his theories). Recently, it has once again inspired the work of LeDoux (2012), which led to the introduction of the survival circuit theory, whose core ideas can be found in Sub-Section 2.6.2. Nowadays, the primary methods of investigating emotions in the animal's brain are rooted in conditioning (both classical and instrumental). Not only does conditioning rely on this idea of survival being improved by affect, it also exploits the associative mechanism by which new triggers for innate emotions are acquired. Furthermore, according to LeDoux (2012), it can be interpreted in light of yet another debate concerning emotions: whether they are inherited from our animal ancestors or learned through countless experiences, usually summarized as '*nature versus nurture*' (Damasio, 1997, 2008; Hawkins & Blakeslee, 2007; LeDoux, 1995; Newell, 1973; J. a. Russell, 2003). For conditioning to be possible, it seems emotions have to be both at the same time. Innate emotions, therefore, are inherited from our ancestors (where ancestors in this case means both our parents, and our animal ancestors), while our everyday experiences teach us how to use them to our advantage.

Empirical results from classical conditioning, as well as functional imaging studies, using both positive and negative stimuli, have identified the amygdala as being the brain area of main interest (Whalen & Phelps, 2009). Its activation for both types of stimuli has led to its inclusion in models and theories of emotions, most notably the limbic system (MacLean, 1952, 1970) and the '*somatic marker hypothesis*' (Damasio, 2008). There its role is usually assumed to simply be '*processing emotions*', without giving too many details about its inner workings or the specific emotional features being processed. Conditioning, on the other hand, has managed to dive deeper in the amygdala's anatomy and explain how associative learning is implemented, as well as the type of information used for processing (Armony et al., 1995; LeDoux, 1990, 1992, 1995, 2014). At an abstract level its anatomy has been compared to a funnel (Whalen & Phelps, 2009), where input stimuli are gradually integrated through the layers that make up the amygdala. The main sensory input to the amygdala being the thalamus, it means that the first few stages integrate data related to the state of the environment. Later inputs to the basal nucleus from associative cortical areas, help contextualize this information (LeDoux, 1990, 1992, 1995). Therefore, the last stage of the process, the central nucleus, is thought to deal with general concepts describing the state of both the environment and the body (via afferent connections from the hypothalamus). Furthermore, the amygdala is also ideally placed to fulfill any survival function attributed to emotions, since it is bi-directionally connected with the hypothalamus and the thalamus. Hence, the central nucleus can influence, via inhibitory projections, the functioning of the autonomic nervous system and the different thalamo-cortical loops responsible for motor and cognitive actions. Another interpretation for the thalamus-amygdala-thalamus loop, is that of a primary attention mechanism. Following that point of view, the central nucleus of the amygdala would then control the amount of information that enters the amygdala itself, as well as other cortices, by increasing or releasing its hold over the thalamus.

It has long been accepted that the hypothalamus is the control center of the autonomic nervous system. However, on the contrary to what was previously believed, the hypothalamus fulfills its role of maintaining the body's homeostasis, not by explicitly controlling other parts of the brain or the body, but by influencing them (Saper & Lowell, 2014). The hypothalamus spreads its influence via the release of hormones, which travel within the brain and cross the blood-brain barrier. Each hormone triggers specific brain areas or organs, which in turn work to re-establish the body's delicate balance. Even though, the hypothalamus has received less attention than the amygdala and its anatomy is not as well-defined, it is safe to assume that a similar funneling phenomenon happens within it. This idea is all the more reinforced by the fact that the accepted role of the hypothalamus is to maintain the body's homeostasis. Hence, monitoring and detecting relevant internal stimuli, are integral parts of its task.

Bringing it all together, the lateral and basal nuclei of the amygdala receive external sensory inputs from the thalamus. They channel these stimuli toward the central nucleus, all the while reducing their dimensionality. The nuclei making up the hypothalamus apply a similar process to internal sensory inputs. Finally, the central nucleus of the amygdala integrates information from both the basal nucleus and the hypothalamus. Thus, making it the ideal brain area for detecting relevant stimuli and trigger an emotional episode if necessary. Consequently, if the overall goal of this thesis, as stated in Chapter 1, is to investigate the role of emotions in virtual agents, a reasonable first step would be to explore the applicability of both LeDoux (2012)'s theory and the results from classical conditioning to those virtual agents. Therefore, the hypotheses, this chapter seeks to investigate, are:

**H**1**-**1 The circuit, described so far and identified as essential to an animal's survival, can also be used by virtual agents to influence their actions and increase their own survival capabilities.

**H**1**-**2 The amygdala acts as the trigger for this survival circuit.

**H**1**-**3 However, This survival circuit is not enough to elicit any emotion, where emotions are differentiated from other mental states by their level of '*arousal*' and '*valence*' (see Sub-Section 2.7.3 for the exact definition of emotion adopted in this thesis).

## 5.1 ProtoEmo: A model based on primitive circuits

To explore the validity and consequences of the hypothesis formulated above, the **ProtoEmo** architecture (depicted in Figure 5.1) has been built. It is a structure made of three main neural populations, each inspired by one of the brain areas mentioned so far: the thalamus, the hypothalamus and the amygdala. **ProtoEmo** has two main input layers. The thalamus receives sensory data describing the state of the environment (external stimuli), while the lateral hypothalamic area processes internal stimuli. The main output layer is the central nucleus of the amygdala, whose activity influences the decision-making process, as well as feeds back to both input layers.

The neural network representing the hypothalamus has been split into two layers. The aforementioned lateral hypothalamic area and the rest of the hypothalamus. Since the dimensionality of the internal inputs, in the case of the following simulations, is quite small

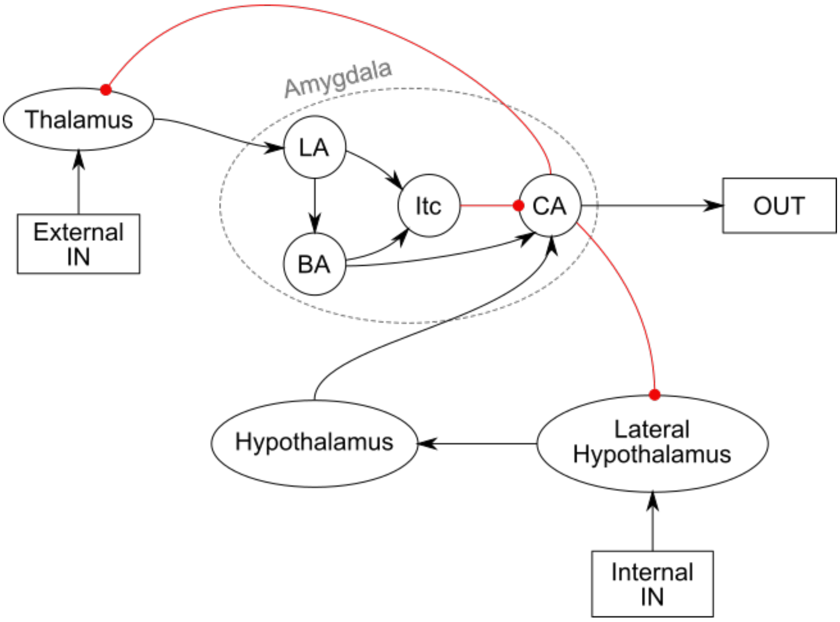


Figure 5.1 *This schema represents the different brain areas, along with their connectomes, that are included in the* ***ProtoEmo*** *architecture. The thalamus and lateral hypothalamic area (LHA) receive sensory input, describing the external environment and the internal state, respectively. The LHA compresses its input and sends it to the hypothalamus, hypothesized to be in charge of informing the central nucleus of the amygdala (CA) of relevant stimuli pertaining to the body's homeostasis. Sensory data concerning the external environment, coming from the thalamus, are forwarded to the lateral nucleus of the amygdala (LA). Within the amygdala, the pattern formed by the inter-nuclei connections has been likened to that of a funnel, by Whalen and Phelps (2009). As is the case for the LHA, this funnel serves to integrate sensory information into more complex concepts. Finally, the CA is hypothesized to warn the rest of the brain about stimuli relevant to the agent's or individual's survival. Via its inhibitory connections, to the thalamus and hypothalamus, the CA implements two feedback control systems. On the contrary to the roles usually assigned to the CA and hypothalamus, they do not directly control the body or other parts of the brain, but simply influence them via a release of different hormones.*

already a single layer is enough to reduce it further. Thus, the lateral hypothalamic area extracts higher-level descriptions of the internal state, which are to be categorized by the hypothalamic layer. Aside from the internal sensory inputs, the lateral hypothalamic area also receives inhibitory projections from the central nucleus of the amygdala. This layer projects directly to the second hypothalamic layer. The second layer, labeled '*hypothalamus*' for simplicity, is then in charge of detecting stimuli relevant to maintaining the body's homeostasis. It is bidirectionally connected with the basal nucleus of the amygdala and sends efferent projections to its central nucleus. Therefore, informing the central nucleus of the presence of any relevant internal stimuli. This first circuit implements one of the low-level control systems.

The thalamus is represented by a single layer, which receives external sensory inputs as well as inhibitory projections from the central nucleus of the amygdala. Its output activity is sent directly to the lateral nucleus of the amygdala. Since the **ProtoEmo** architecture does not contain any neocortical layer yet, the thalamus-amygdala-thalamus loop is interpreted here as a primary attention mechanism. Through this second low-level control system, the amygdala can restrict the amount and type of information that enters its lateral nucleus by further inhibiting or releasing its grasp on the thalamic layer.

Finally, the amygdala has been split into its four main components: the lateral, the basal, the intercalated and the central nuclei. As briefly described in the introduction, the organization of the lateral, basal, and central nuclei resembles a funnel. As a matter of fact, it would make more sense to compare the flow of information inside the amygdala to that of the visual system. The lateral nucleus is usually considered as the interface to the amygdala (Whalen & Phelps, 2009), receiving sensory information related to the state of the world from the thalamus, as well as the somatosensory cortex (see Chapter 4 for a description of the '*dual pathways*' between the thalamus and the amygdala). As the lowest level area in the amygdala, the lateral nucleus can only detect and encode for very simple features of the world. This is similar to how neurons in the first level of the visual system are tuned to detect edges in a specific orientation. The lateral nucleus then projects directly to the basal and the intercalated nuclei. Continuing the metaphor with the visual system, based on the activation patterns of the lateral nucleus the basal nucleus can detect more complex features. The final stage of the amygdala, the central nucleus, receives afferent connections from both the basal and intercalated nuclei. As it is the highest level of the amygdala, the central nucleus is, therefore, capable of representing intricate features relevant to the survival of the individual. Consequently, the amygdala's central nucleus can be conceived as a group of detectors, an idea not unlike MacLean's (1949) '*cerebral analyzers*' or Scherer and Ekman's (2010; 1984) '*Stimulus Evaluation Checks*' (introduced in Sub-Section 2.4.1). The intensity of each detector's output is here interpreted as a measure of the salience of the corresponding feature. As a result, the activation pattern of the amygdala's central nucleus represents the saliences of all the features from the environment that are of major concern for the survival of the individual (an idea supported by McClure et al. (2004, p. 262)). As described in Chapter 4, Armony et al. (1995) showed that through the interplay of Hebbian learning and lateral inhibition it is plausible for the amygdala's neurons to develop those detectors, as well as tune them via conditioning (as explained in Appendix A lateral conditioning has been implemented in the **ProtoEmo** controller, however, Hebbian learning has been replaced with a genetic algorithm).

In the brain of animals, the central nucleus sends inhibitory feedback connections to both the thalamus and lateral hypothalamic area. Additionally, for the purpose of the simulation described below the central nucleus also projects to an output layer, whose activity is then used in the decision-making process (see Appendix A for details on the policy followed by the **ProtoEmo** controller). Therefore, as described earlier, neither central nucleus nor hypothalamus directly control motor and cognitive actions, instead they influence them via release of different hormones.

From the activity received from the basal and lateral nuclei, the intercalated layer predicts the amount by which the central nucleus should be inhibited. Therefore, it has been hypothesized to implement a basic habituation mechanism, by Whalen and Phelps (2009) and O'Reilly and Munakata (2013), since its inputs are similar to that of the central nucleus, while its learning rate is slightly slower. Section A.1 provides more details as to the actual implementation of the brain areas described so far, as well as the projection patterns between them.

## 5.2 Testing ProtoEmo on a resource foraging task

To test the performances of the **ProtoEmo** architecture, an adapted version of the '*one-resource foraging task*', described by Scheutz (2004), was used. A standard resource foraging task usually consists in having a group of agents survive within a dynamic environment. To fulfill their goal, the agents must explore the environment, in search of energy resources to forage. Since undertaking any action expends some amount of energy, this experiment is designed to test the strategy implemented by the group of agents to share the available resources among its members. The results Scheutz (2004) gathered via his implementation of the resource foraging task, allowed him to conclude that virtual agents endowed with the concept of emotions are '*better*' at surviving, than asocial and non-adaptive agents. Where '*better*' was defined as a measure of the number of agents remaining at the end of a simulation. Since Scheutz (2004) has already gone to great lengths to show that his emotional agents have better survival skills, our goal is to compare their performance with a group of agents controlled by the **PrimEmo** architecture. To this end, we implemented our own version of the resource foraging task.

### 5.2.1 Experimental setup

The environment for this experiment is a two-dimensional surface of infinite proportions. This simply means that wherever an agent decides to go, it will never '*knock*' against any virtual edge. Around the origin of this surface, an area of 1800 in width and 1800 in length is virtually defined. It serves to restrict the possible locations at which energy resources can appear and therefore the amount of movement agents have to perform between foraging events. At the beginning of every simulation, circular obstacles are randomly scattered inside the pre-determined area. These obstacles, of radius 0.05, are solid disks that agents have to avoid while traveling between resources. They cannot be broken and remain in the same position for the entire duration of the simulation. Similarly, resources are represented by circles (also of radius 0.05) that can be entered by any agent wishing to forage its energy. By default each resource holds 800 units of energy, and disappears whenever it has been emptied. Regardless of the foraging rate only one new resource appears, at a random location within the pre-determined area, at the end of each simulation cycle.

Instead of being simple points in space, as is the case in Scheutz's implementation, agents are represented by disks of radius 0.037. Furthermore, each agent is characterized by a level of energy (set initially to 2000 units) and an '*action tendency*', which is a coefficient in the range $[0, 1]$ indicative of the probability the agent chooses to fight rather than flee. During each simulation cycle, agents have to decide which action to perform next, based on the state of the environment and their own internal state. To make matters worse for the agents, a '*Fog of War*' has also been implemented. This translates into a sensory horizon beyond which agents are not able to perceive the state of the world. For this specific experiment the range within which sensors where working, was set to a 300 units radius around the agent. Hence, based on these incomplete sensory data agents have the chose between the following five actions:

1. Move: this is the default action if nothing appears in the agent's vicinity. The agent determines the speed and direction of its movement using the equation:

   $$D = \sum_n g_r \times resource(n) + \sum_m g_a \times agent(m) + \sum_k g_t \times obstacle(k) \qquad (5.1)$$

   Where $g_r$ is a coefficient indicative of the agent's need in energy, $g_a$ represents how much an agent wants to enter a battle (if positive the agent will be moving toward others, while a negative value means that the agent will actively avoid others), and $g_t$ corresponds to obstacle avoidance (similarly, a positive value means the agent will try as much as possible to run into obstacle, while a negative value forces it to avoid obstacles). $resource(n)$, $agent(m)$ and $obstacle(k)$ are the scaled vector from the agent's current position to the $n^{th}$ resource, $m^{th}$ nearby agent (excluding itself) and $k^{th}$ obstacle, respectively.
   The maximum speed at which an agent can move is set to 4. Additionally, if its level of energy is below the critical level (set to 400 units), its speed is further restricted to a maximum of 1, to avoid using the remainder of its resources on movement alone. While moving, an agent expands energy at a rate equal to the square of its linear speed.

2. Fight or Flee: when two or more agents are within a radius of $fight\_thres = 0.3$ of each other, a conflict situation arises. The agent has to decide whether to fight or run away. The process by which this decision is made, depends on the type of agent (as detailed in the next Sub-section). However, it is always based on the agent's current action tendency. On the one hand, if it decides to fight, an agent looses 50 units of energy over a single cycle. On the other, fleeing is equivalent to running away from the closest enemy, at a speed of 7 for a random number of cycles (chosen in the range $[5, 10]$).

3. Forage: when the agent is on top of a resource it will consume its energy, until empty. There is no limit to how much energy an agent can consume.

4. Procreate: if an agent manages to survive for more than 250 cycles and has more than 2200 units of energy left, it will then automatically procreate. This means that a new agent will appear at a random location inside the pre-determined area. The act of procreation costs 2000 units to the parent.

Finally, at the end of a simulation cycle each agent loses one unit of energy as processing power, urging them to move around and forage for energy as much as possible, since an energy level of zero is equivalent to death. Upon dying an agent simply disappears from the environment.

### 5.2.2 Scheutz's emotional agents

In Scheutz's experiment four types of agent were defined: asocial non-adaptive, social non-adaptive, asocial adaptive and emotional (social and adaptive). When a social agent finds itself in a conflict, it fights only if its action tendency is the highest, otherwise it runs away. Thus, battle between social agents are resolved after one turn only. Asocial agents, in the same situation, do not take the action tendency of others into consideration, but use their own action tendency as a probability for fighting, rather than fleeing. Therefore, a conflict between asocial robots can last until one of them decides to run away or dies.

The capacity of an agent to adjust its action tendency makes the difference between adaptive and non-adaptive individuals. On the one hand, if a robot is non-adaptive then its action tendency is fixed and chosen at random upon initialization, following a Gaussian distribution of mean 0.5 and a standard deviation of 0.125. Adaptive agents, on the other hand, adjust their action tendency depending on whether they lost or won the last fight. A losing agent increases its action tendency, thus increasing its probability of winning the next battle, while a winning agent decreases its action tendency. The adaptation rule (AR), as described by Scheutz (2004), defines the amount by which the agent's action tendency should be increased or decreased. If $r$ is the basic action tendency (randomly initialized at the beginning of the simulation) and $m$ is the current action tendency, then:

- After a loss:

$$AR^{+}(m) = \begin{cases} m + \frac{(1-m)}{2} & \text{if } m \geq r \\ 2 \times m & \text{if } m \leq \frac{r}{2} \\ r + (2 \times m - r) \times \frac{(1-r)}{2 \times r^2} & \text{otherwise.} \end{cases} \tag{5.2}$$

- After winning:

$$AR^{-}(m) = \begin{cases} 2 \times m - 1 & \text{if } m \geq r + \frac{1-r}{2} \\ \frac{m}{2} & \text{if } m \leq r \\ \frac{r}{2} + \frac{r \times (m-r)}{(1-r)^3} & \text{otherwise.} \end{cases} \tag{5.3}$$

Emotional agents have the added possibility of setting a non-zero value for their fighting tendency $g_a$, by following the equation:

$$g_a = 100 \times action\_tendency - 50 \tag{5.4}$$

Regardless of the agent's type, the initial value for the other gains are fixed to $g_r = 20$ and $g_t = -20$.

### 5.2.3 Control based on the ProtoEmo architecture

To evaluate the survival capabilities the **ProtoEmo** architecture is hypothesized to confer to artificial agents, it has been implemented as an artificial neural network. Each of the brain areas detailed in Section 5.1 is represented by its own neural population. The only exception being the amygdala, whose nuclei are separated into independent layers, rather than being part of a single amygdala network. This is to allow a more precise control over the underlying mechanisms of the amygdala. The basic building blocs for this architecture are normalized leaky rectified linear units ('*leaky ReLU*', first described by Maas, Hannun, and Ng (2013)). More details about the actual implementation of the **ProtoEmo** controller are available in Appendix A. To simplify the programing process and take full advantage of the acceleration provided by GPU computing, the python libraries Keras and Tensorflow were used.

As mentioned previously, the **ProtoEmo** architecture relies on two sets of inputs. The external inputs, gathered by the agent's range sensors and through communication with other nearby agents, are first normalized before being sent to the thalamus. To perform its resource foraging task each agent uses the following information to determine the state of the environment:

- The maximum action tendency: This value is sampled over all the nearby agents. It helps the current agent to decide when to fight and when to flee. In the case, where there are no agents in its vicinity it defaults to 0.
- The distance to the closest obstacle: The distance to the closest obstacle is computed relative to the current position of the agent. If there are no obstacles within the sensory horizon, then the distance is set to 0 by default. This way the neurons encoding for this value will not participate in the activation of the thalamic population.
- The distance to the closest source: The shortest distance to an energy source is computed the same way as the one for the obstacle.
- The distance to the closest agent: The same remains true for this last distance, which is relative to the agent's current position. Moreover, it allows the agent to avoid any conflict with hungrier agents. Thus, preventing any energy depletion from running away or fighting.

The hypothalamus being the control center for the autonomic nervous system, it receives information from the internal inputs via its lateral area. In a manner similar to the external inputs, the internal ones are also normalized before being sent to the corresponding input layer. To perform its homeostatic duties, the hypothalamus monitors the following variables:

- The energy level: Same as for any other creature, the agent has to watch over its level of remaining energy, since depletion means death.
- The current action tendency: In this specific case the agent's action tendency, can be interpreted as indicating how desperate, a particular agent is, to forage the nearest energy source. It could be likened to what is usually described as '*motivation*' in animals.
- The level of danger: Danger is a variable computed based on the distance between the closest agent and the closest obstacle (see Appendix A for a definition of this variable). It was introduced early on to elicit fear in the agent.

Given the description of **ProtoEmo** at both the architectural (see Section 5.1) and controller levels, a few improvements over Scheutz's (2004) implementation of his emotional agents stand out. To begin with, agents controlled by the **ProtoEmo** architecture can use the output of the network to influence the values of all their tendencies (more details relating to this process are available in Sub-Section A.1.5), not just their action and fight tendencies. Therefore, if for example an agent finds itself in a perilous situation, then it can suppress its hunger to move away from danger before resuming its search for an energy source. On the contrary, Scheut's emotional agents have a fixed foraging tendency, which implies that they are always '*hungry*' and will seek energy in priority regardless of the situation. Since **ProtoEmo** has been implemented in populations of artificial neurons, it is highly parallelizable and can theoretically approximate any function by combining the output of its different non-linear transforms. As a result of its complexity and the free parameters available to the **ProtoEmo** controller it can devise more complex and potentially better strategies. Even though, this can be conceived as an improvement over Scheutz's emotional agents, the next section discusses how in the context of the resource foraging task learning an optimal strategy does not necessitate an elaborate architecture. In fact, only a few modifications to the adaptation rule and fighting tendency defined by Scheutz (2004) (in Equations 5.3— 5.2 and Equation 5.4 respectively) are sufficient to replicate the strategy adopted by the **ProtoEmo** controller. A further consequence of the inherently parallel nature of the **ProtoEmo** controller is that it can provide values for all four tendencies at the same time, rather than rely on different functions to compute each one sequentially, hence, reducing the time needed for an agent to decide on which action to perform next. Given the simplicity of the simulation and the few rules defined by Scheutz (2004), the difference in responsiveness might not seem worth the overhead. However, any increase in task complexity will emphasize the gain in simulation time.

Aside from the benefits afforded by its underlying neural network, each agent controlled by **ProtoEmo** is aware of its own body in addition to the state of the world. Although, **ProtoEmo** has access to the same input features as Scheutz's emotional agents, **ProtoEmo** actually makes use of all available inputs to compute an agent's tendencies (as described in more details in Appendix A). On the contrary, as made apparent by Equation 5.3, Equation 5.2, and Equation 5.4 the only parameters in the adaptation rule are the agent's original and current action tendencies, as well as the result of the last conflict (victory or defeat). Moreover, owing to **ProtoEmo**'s organization in a hierarchy of interconnected layers simulating the brain and its functional layout, sensory information describing the world and bodily state is not processed all at once. Instead, it is split into two separate streams (a fact reminiscent of Papez's (1937) circuit described in Section 2.6). In each of those streams, the further the data travels through the hierarchy the more it is consolidated into higher-level concepts, until it reaches the central nucleus of the amygdala and is categorized as being relevant to the survival of the individual, or not. This is a perspective that is briefly mentioned above and further explored in Appendix A.

Since the **ProtoEmo** architecture has many recurrent connections (see Figure 5.1), for each new set of inputs the network was run until stability. Where stability is defined as the activation values of the output neurons being the same for two consecutive activation cycles. The output layer is made of four neurons. The first three corresponding to the coefficients ($g_r$, $g_a$ and $g_t$), which influence the direction and speed of the agent's movement (see Equation 5.1). The fourth neuron represents the value of the current action tendency. Having the values of the output layer contribute to both movement and motivation, fits with the idea that the amygdala and hypothalamus implement control systems that only influence action selection, but do not make hard decisions (see Sub-Section A.1.5 for more details on how those values are used to influence the decision-making process).

A recurrent idea within affective science is that emotions evolved and are therefore inherited from our ancestors. Furthermore, because it has been hypothesized that emotions facilitate survival, learning in-between simulations is achieved by a genetic algorithm (details of its implementation are described in Section A.2). The simple genetic algorithm used in this experiment tries to optimize for the size of the population remaining at the end of each simulation, as

Table 5.1 This table details the statistics describing the number of survivor at the end of a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 89.49 | 100 | 24.415 | 94.275 | 84.705 |
| Scheutz | 9.10 | 100 | 0.927 | 9.282 | 8.918 |

Table 5.2 This table details the statistics describing the maximum size the group of agents reached during a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 89.49 | 100 | 24.415 | 94.275 | 84.705 |
| Scheutz | 10.03 | 100 | 0.171 | 10.064 | 9.996 |

well as the maximum size reached by a population, at any point in time. The second objective was introduced in an effort to promote equity within the population.

## 5.3 Results

To collect the data necessary to answer the different hypotheses laid out in Section 5.1, the controllers described in Sub-Section 5.2.2 and Sub-Section 5.2.3 went through two phases of development. The first was the '*learning*' phase, in which the controller based on the **ProtoEmo** architecture used its genetic algorithm to optimize its strategy for 70 generations. Since there is no learning involved in Scheutz's emotional controller, no changes were made in this first phase. During the second '*testing*' phase, both controllers went through 100 simulations, each lasting 2500 turns, where a simulation cycle is characterized by all agents having performed an action and the environment being updated. For both phases, each simulation was initialized with a group of ten agents, each agent having an initial level of energy of 2000 units. Furthermore, at the beginning of each foraging task Scheutz's emotional agents had their tendencies randomly sampled from a normal distribution with mean 0.5 and standard deviation 0.125. In contrast, Agents using the **ProtoEmo** controller had their tendencies computed directly based on the neural activity of the architecture's output units.

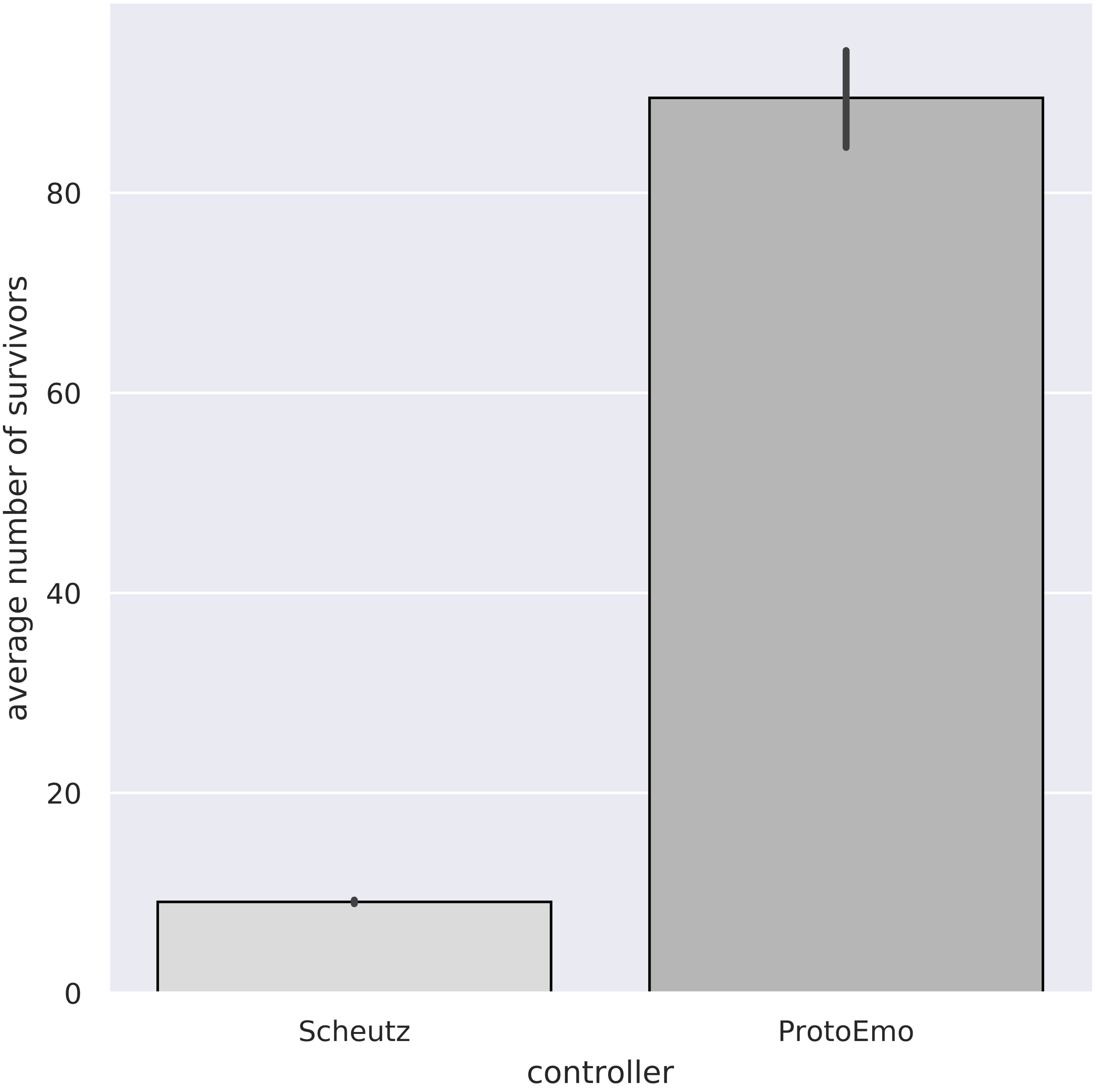


Figure 5.2 *Bar plot representing the average number of survivors for each type of controller. Survivors are simply defined as the agents left alive once the simulation reaches its maximum number of cycles. The error bar represent the* 95% *confidence interval.*

Table 5.3 This table details the statistics describing the life span of agents within a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 2461.851 | 9241 | 232.279 | 2466.587 | 2457.115 |
| Scheutz | 2366.903 | 1004 | 451.503 | 2394.832 | 2338.975 |

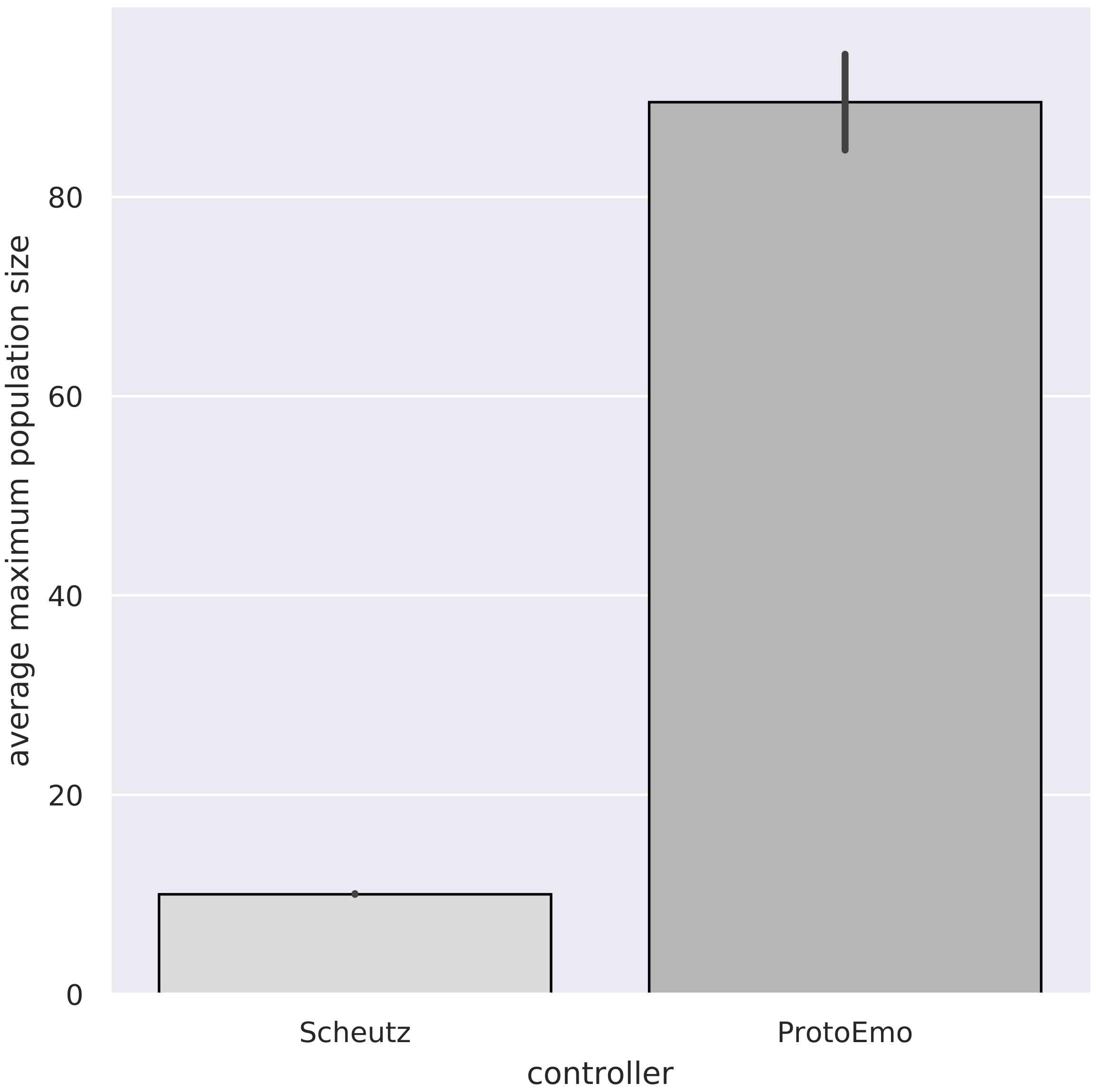


Figure 5.3 *Diagram depicting the average size of the maximum population, per game and per controller. It is different from the number of survivors, in that some of the agents, present in this population, might die before the end of the game, without being replaced. The error bar represent the* 95% *confidence interval.*

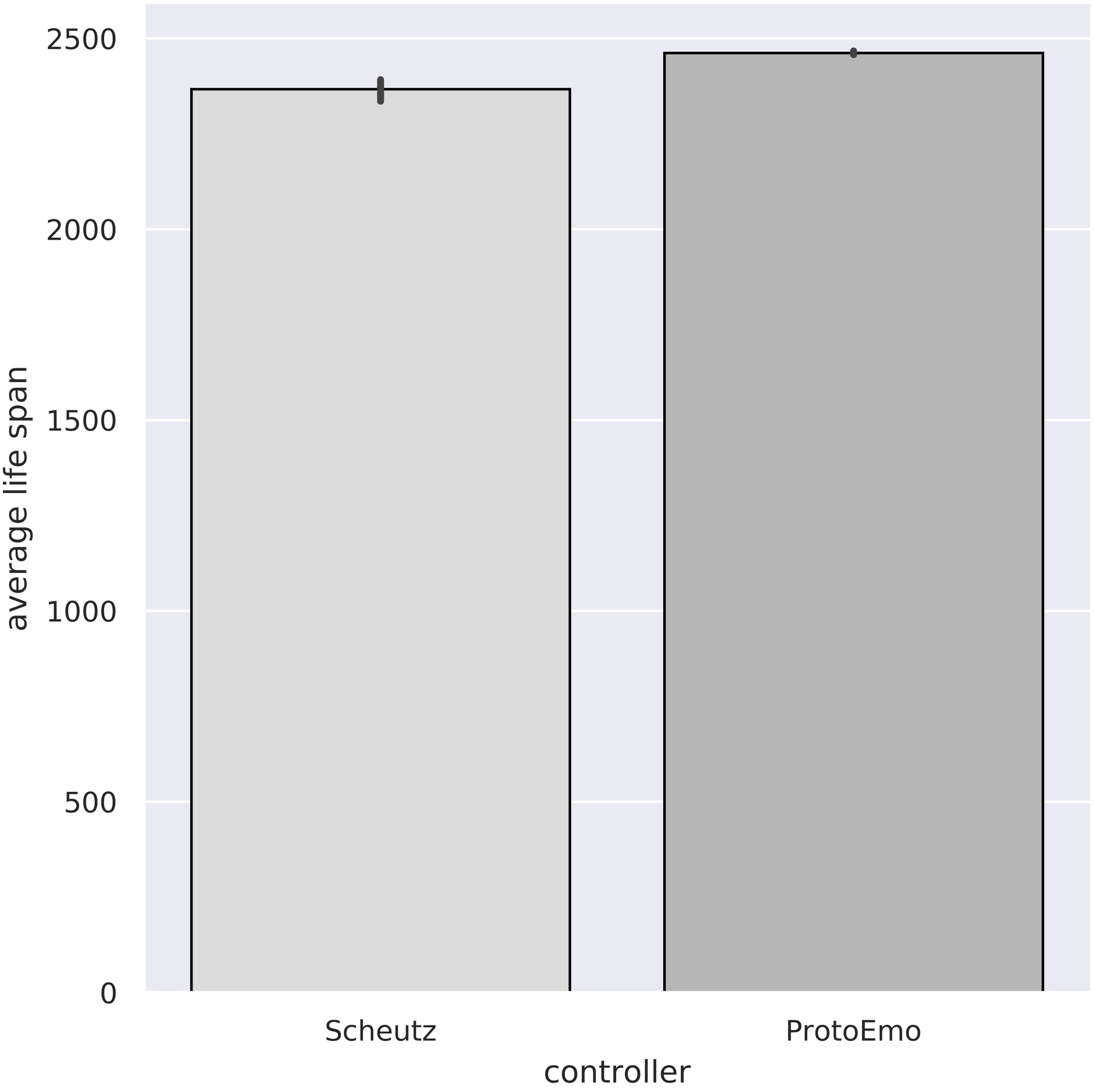


Figure 5.4 *Bar chart representing the average life span of an agent, in number of cycles, for each type of controller. The error bar represent the* 95% *confidence interval.*

Table 5.4 This table details the statistics describing the percentage of time agents spend procreating during a simulation.

| | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ctrl | | | | | |
| ProtoEmo | 1.559e-03 | 9241 | 2.433e-03 | 1.609e-03 | 1.509e-03 |
| Scheutz | 2.831e-06 | 1004 | 5.620e-05 | 6.307e-06 | -6.458e-07 |

Table 5.5 This table details the statistics describing the percentage of time agents spend fighting during a simulation.

| | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ctrl | | | | | |
| ProtoEmo | 1.549e-06 | 9241 | 1.178e-04 | 3.952e-06 | -8.533e-07 |
| Scheutz | 1.857e-04 | 1004 | 9.136e-04 | 2.422e-04 | 1.292e-04 |

The bar charts of Figure 5.2, Figure 5.3 and Figure 5.4, show the results for the average number of survivors, the average maximum population and the average life span of an agent, respectively. The error bars for all the graphics represent the 95% confidence interval. From these results it is clear that the circuits identified by LeDoux (2012) as important to the survival of animals can also be used to influence virtual agents and increase their own survival potential (this is referred to as H1-1 in Section 5.1). Indeed, both the average number of survivors and the average maximum population for **ProtoEmo** is almost nine times higher than that of Scheutz's emotional agents (see Tables 5.1 — 5.3 for a more detailed results).

Remarkably, the average life span is quite similar across all controller types. Moreover, taking into account the data from Figure 5.5 and Table 5.4, it is interesting to note that the average time spent reproducing per agent and per game is three orders of magnitude higher for **ProtoEmo** when compared to Scheutz's emotional agents. This discrepancy, and, therefore, the difference in the number of survivors at the end of the game can be explained by the results of Figure 5.9 and Table 5.8. Indeed, Scheutz's emotional agents having a lower energy level on average means that they have to put more effort into exploring the environment and foraging for food before being able to procreate. However, the more time is spent exploring and foraging the higher the probability two or more agents will find themselves in a conflict. This in turn increases the risk of the agent losing more energy or even dying before being able to procreate.

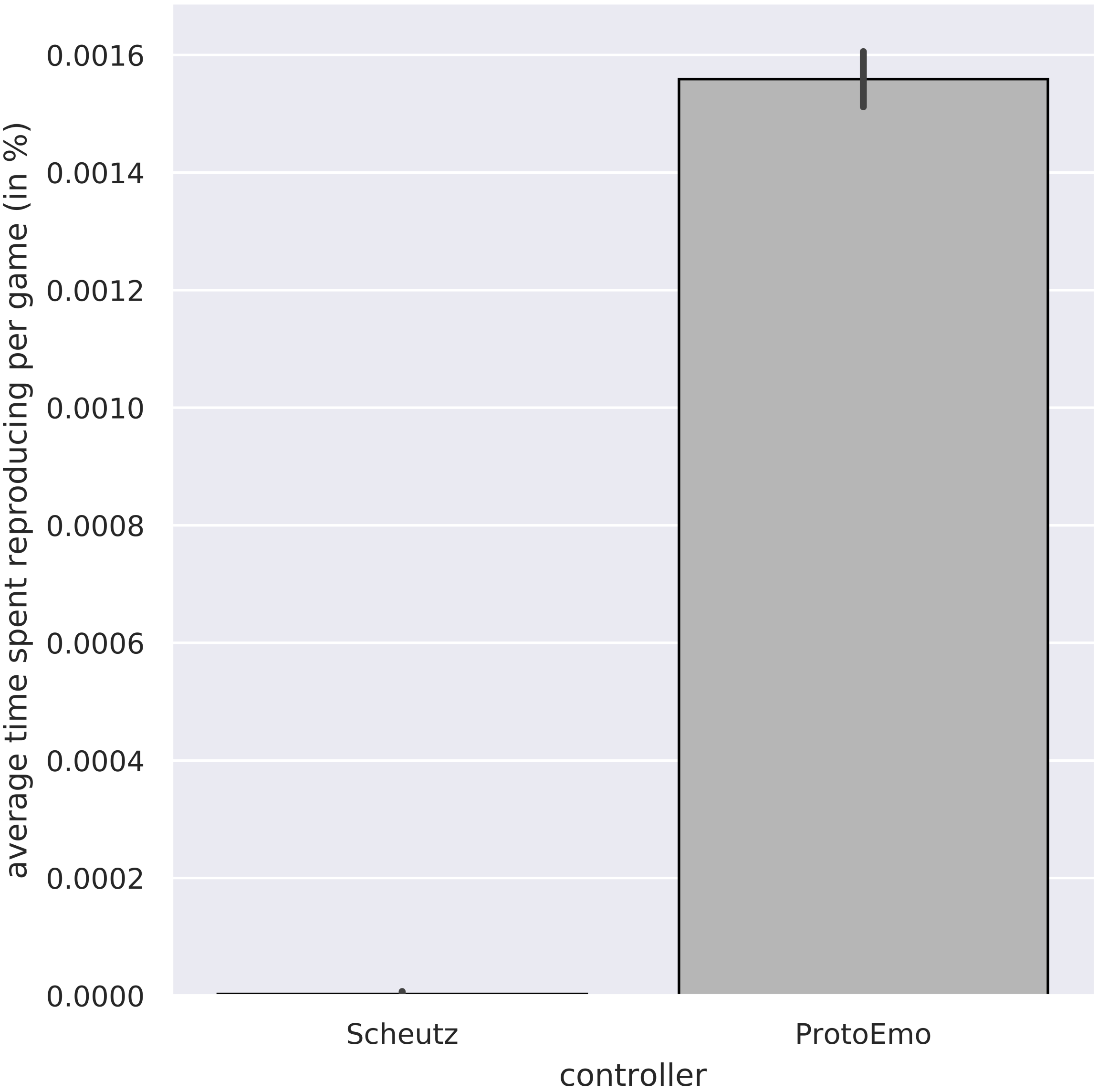


Figure 5.5 *Illustration of the average time an agent spends reproducing, for each controller type, per game. It is expressed as a percentage, since not all agents have the same life span, so the value needs to be normalized. The error bar represent the* 95% *confidence interval.*

Table 5.6 This table details the statistics describing the percentage of time agents spend running away during a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 1.736e-06 | 9241 | 1.669e-04 | 5.139e-06 | -1.667e-06 |
| Scheutz | 1.256e-03 | 1004 | 4.019e-03 | 1.505e-03 | 1.007e-03 |

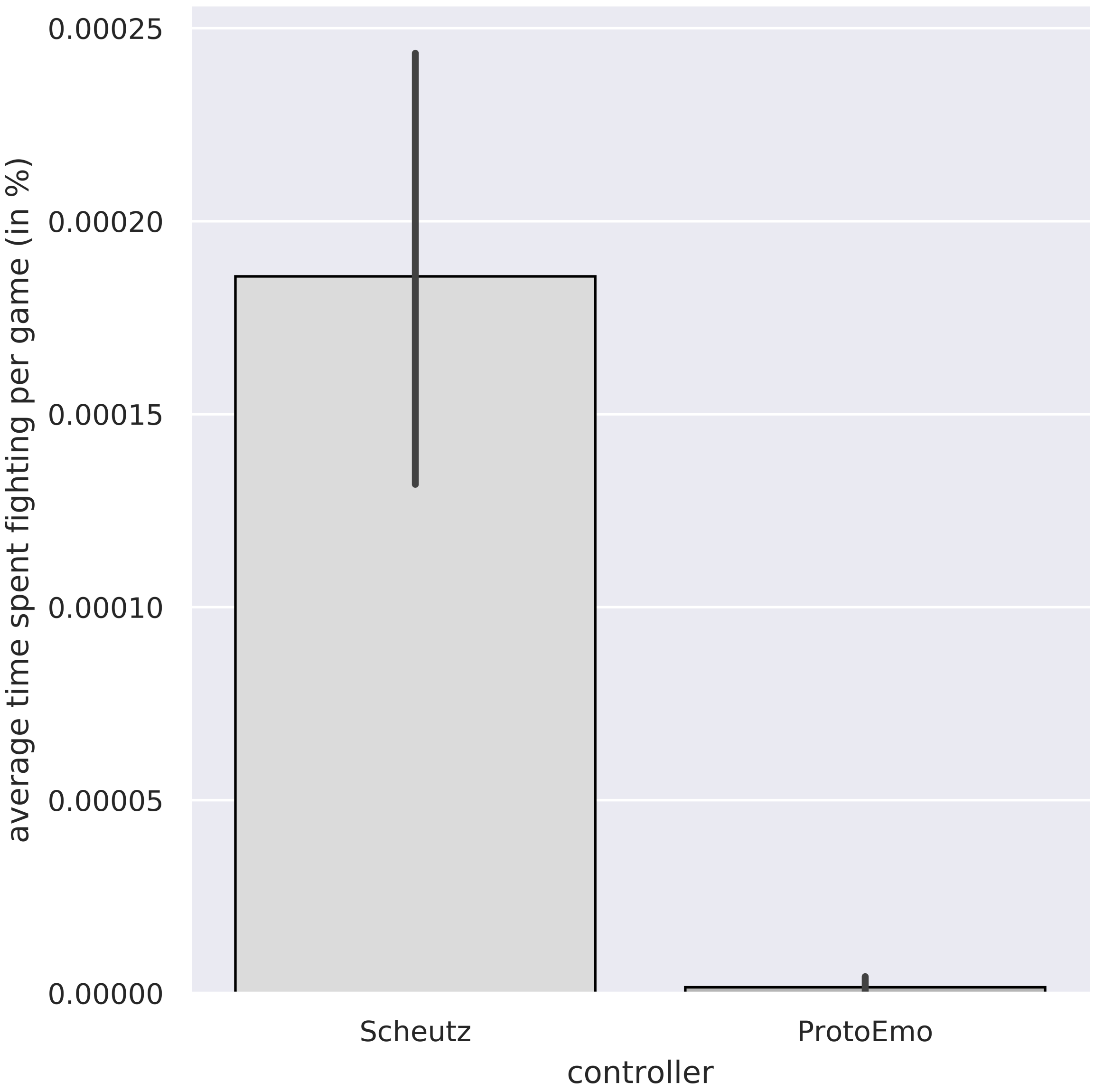


Figure 5.6 *Plot showing the average time spent fighting for each controller type. The error bar represent the* 95% *confidence interval.*

Table 5.7 This table details the statistics describing the percentage of time agents spend foraging during a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 0.005 | 9241 | 0.004 | 0.005 | 0.005 |
| Scheutz | 0.010 | 1004 | 0.002 | 0.010 | 0.010 |

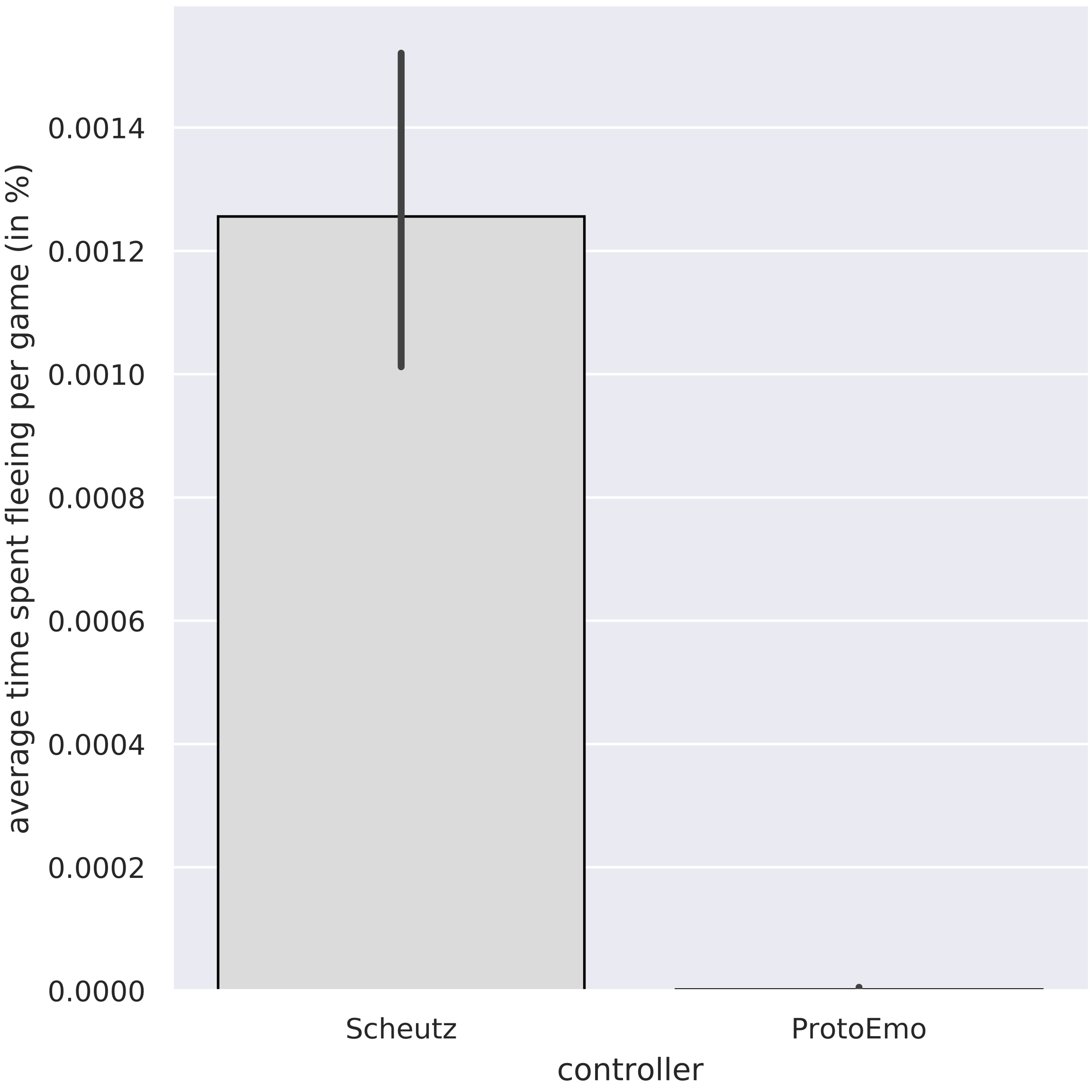


Figure 5.7 *Bar chart depicting the average time an agent spends fleeing during a game, per controller type. The error bar represent the* 95% *confidence interval.*

Table 5.8 This table details the statistics describing the energy level of agents during a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 1296.644 | 7415732 | 526.746 | 1297.024 | 1296.265 |
| Scheutz | 898.768 | 2376205 | 416.519 | 899.298 | 898.239 |

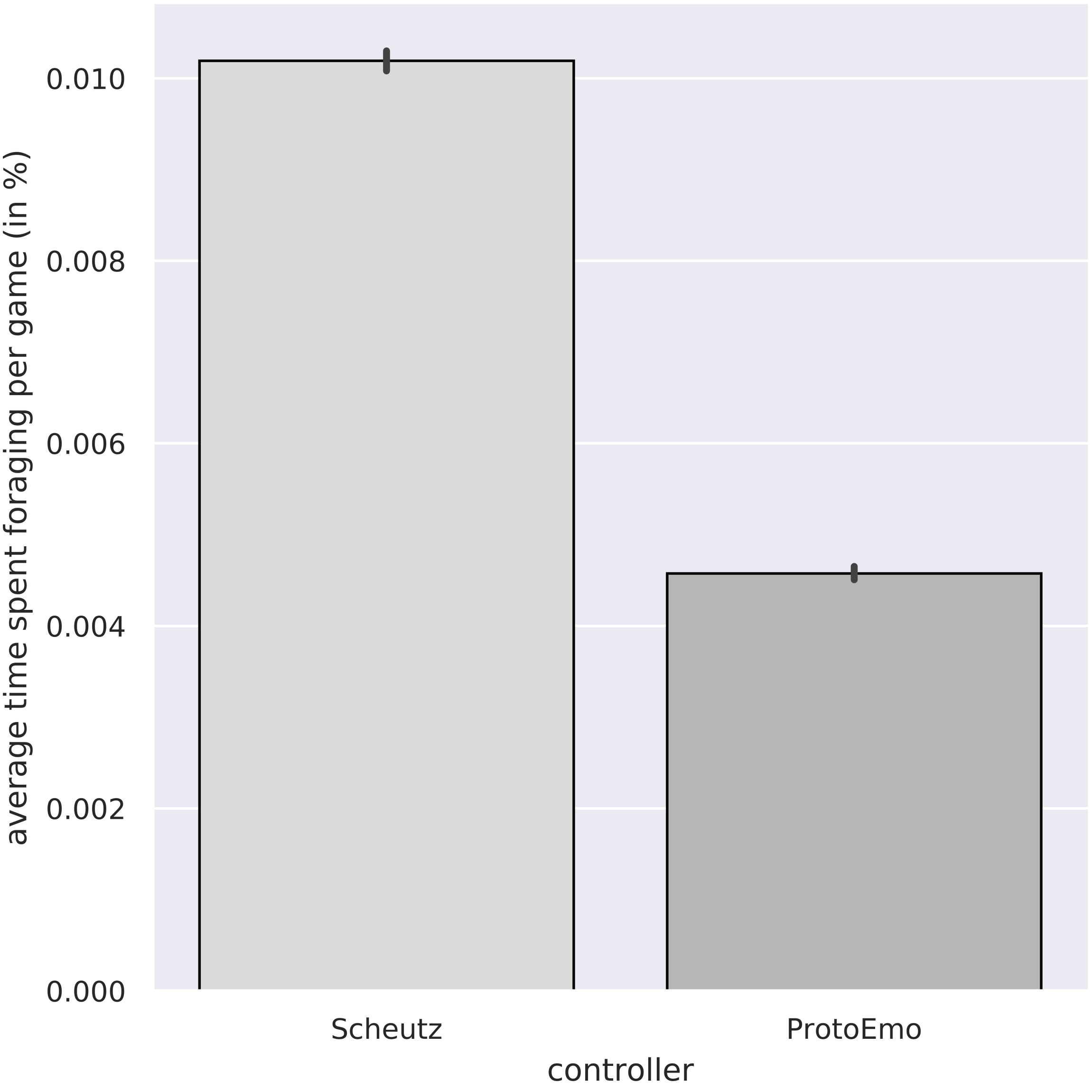


Figure 5.8 *Diagram showing the average time spent foraging for energy, for each controller type. The error bar represent the* 95% *confidence interval.*

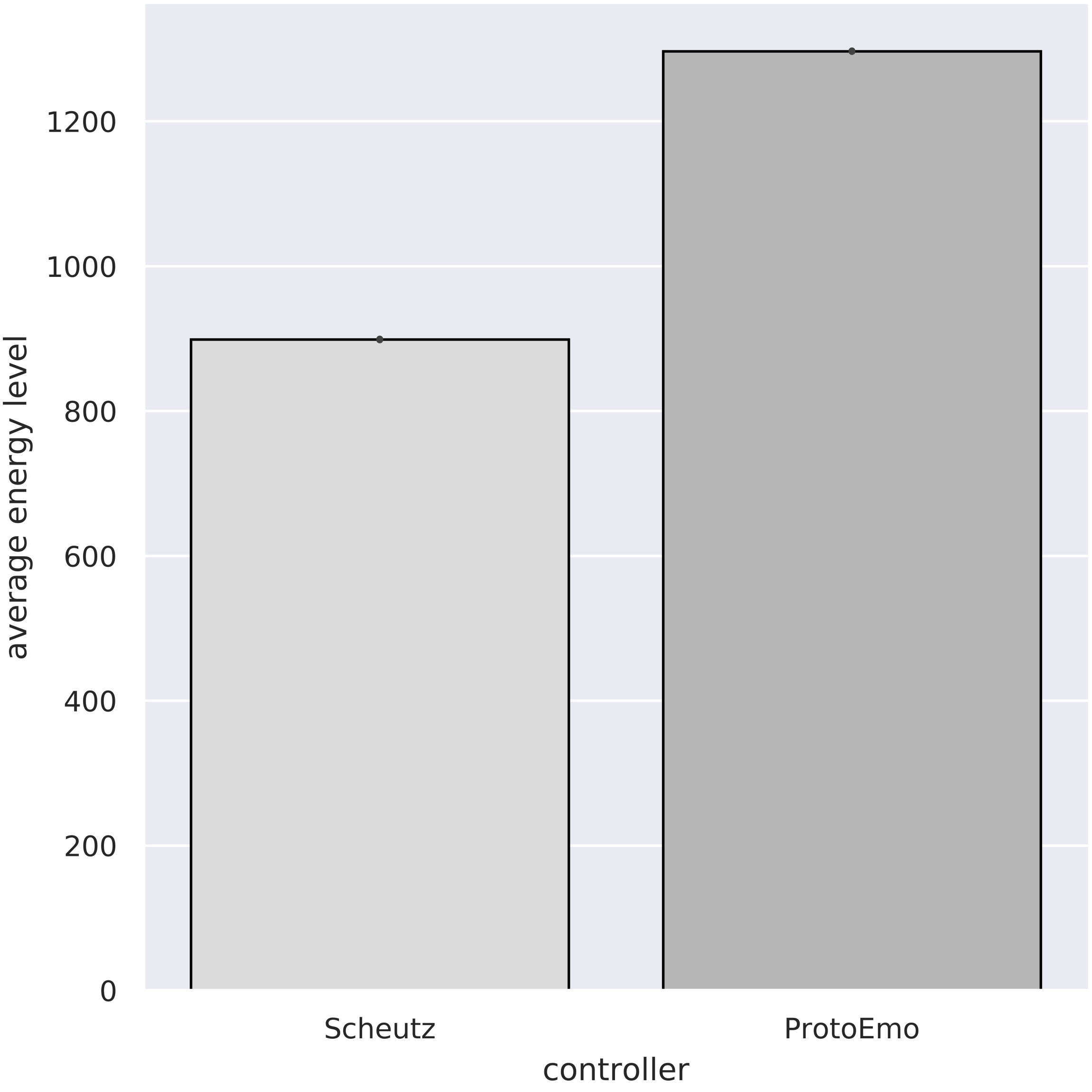


Figure 5.9 *Figure depicting the average level of energy, for each controller type. The error bar represent the* 95% *confidence interval.*

By taking into consideration Figures 5.6 – 5.10, as well as Tables 5.5 — 5.9, it is possible to extrapolate the strategies employed by each type of controller. Scheutz's emotional agents apply a more egoistic strategy, where each agent has to fight for its survival. Indeed, from Figure 5.6 and Figure 5.7 it is clear that they spend more time fighting each other than the agents controlled by **ProtoEmo**. A fact made clear by the detailed results of Table 5.5 and Table 5.6. Fighting and fleeing are two actions which use a great amount of energy, especially for the fleeing agents. As a result, Scheutz's agents need to regularly forage for energy as can be seen in Figure 5.8 and Table 5.7. This in turn leads to more battles and so even more energy used. The main consequence of this strategy, as shown in Figure 5.9 and in Table 5.8, is that the average energy level per agent is only around 900 units, a level lower than **ProtoEmo**'s average of 1300 units and far too low to meet the requirement for procreation (which is 2200 units). Consequently, all Scheutz's emotional agents can do during a simulation is to move toward the closest energy source, battle for it and finally forage its energy to be able to fight for the next source. This strategy leaves no room for any reproduction, since an agent that just procreated would be too weak to either fight or flee, and would simply end up dying.

In contrast, even though **ProtoEmo** achieves a higher population size on average (according to Figure 5.3 and Table 5.2), very few battles take place. Consequently, almost no energy is used on either fleeing or fighting with other agents (as can be seen in Figures 5.6 and 5.7, and Tables 5.5 and 5.6). Instead, agents seem to avoid each other and only move to discover new energy sources. This means that for an agent that just reproduced and gave most of its energy to its offspring, it is quite easy to forage for energy and, thus, avoid death. Against Scheutz's law of the strongest, **ProtoEmo** therefore suggest a 'first come, first served' alternative.

Another possible strategy that the **ProtoEmo** controller might be implementing would involve a few select agents constantly reproducing and foraging, whereas the rest of the population spends most of its time exploring, while avoiding each other and occasionally foraging for energy. This alternate strategy, would help explain why so much of an agent's time is dedicated to simply moving around (according to Figure 5.10 and Table 5.9, this is true for both Scheutz's emotional agents and for **ProtoEmo**). However, a more optimal strategy would have any agent not currently foraging for resources or procreating, stay idle to conserve its energy as much as possible, hence, dedicating all resources to the procreating agents. It is possible for an agent to not move at all by setting its tendencies to zero. This means that on each simulation cycle, the agent would only lose one unit of energy to processing. However, because agents are initialized with only 2000 units of energy and each simulation lasts 2500

Table 5.9 This table details the statistics describing the percentage of time agents spend moving during a simulation.

| ctrl | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ProtoEmo | 0.994 | 9241 | 0.006 | 0.994 | 0.994 |
| Scheutz | 0.988 | 1004 | 0.003 | 0.989 | 0.988 |

cycles (as described in Sub-Section 5.2.1), if an agent decides to remain idle it will still have to forage at least one energy source per simulation. This strategy is not without risks though, since an agent whose energy level is below 400 units has its speed limited to only 1 (as explained in Sub-Section 5.2.1). Therefore, the agent risks dying on the way to foraging the nearest energy source. Consequently, for this strategy to work any agents not selected for procreation, would have to strike a balance between remaining idle to conserve energy with foraging to stay alive and be able to move at full speed to the nearest energy source. It would be interesting to explore which of the strategies is the optimal one, and if given time the genetic algorithm would converge toward this solution.

Although, the strategies employed by Scheutz's emotional agents and the **ProtoEmo** controller seem vastly different, they could be reconciled by replacing the adaptation rule specified by Scheutz (2004) in Equations 5.2 and 5.3. Rather than defining the action tendency based on an agent's victories or defeats, which is only indicative of its '*frustration*', the action tendency should be dependent on the level of energy. Therefore, the lower an agent's energy level is, the more it is willing to fight for an energy source. A similar concept should also be applied to the fighting tendency ($g_a$). An agent's fighting tendency should not only always be negative ($g_a \leq 0$), but also increase the lower the level of energy is. Thus, an agent with a high level of energy will avoid fighting at any cost, whereas an agent with a low energy level while not seeking a fight directly will be less prudent when exploring the environment in search for energy. The following definitions for the action and fighting tendencies could yield a strategy similar to the one adopted by the **ProtoEmo** controller:

$$action\_tendency = \frac{1}{1 + energy\_lvl} \tag{5.5}$$

$$g_a = 50 \times (action\_tendency - 1) \tag{5.6}$$

Considering Figures 5.11 and 5.12, it should be noted that although the tendencies are not equal to the activation of the neural units in the central nucleus of the amygdala (CA), they do

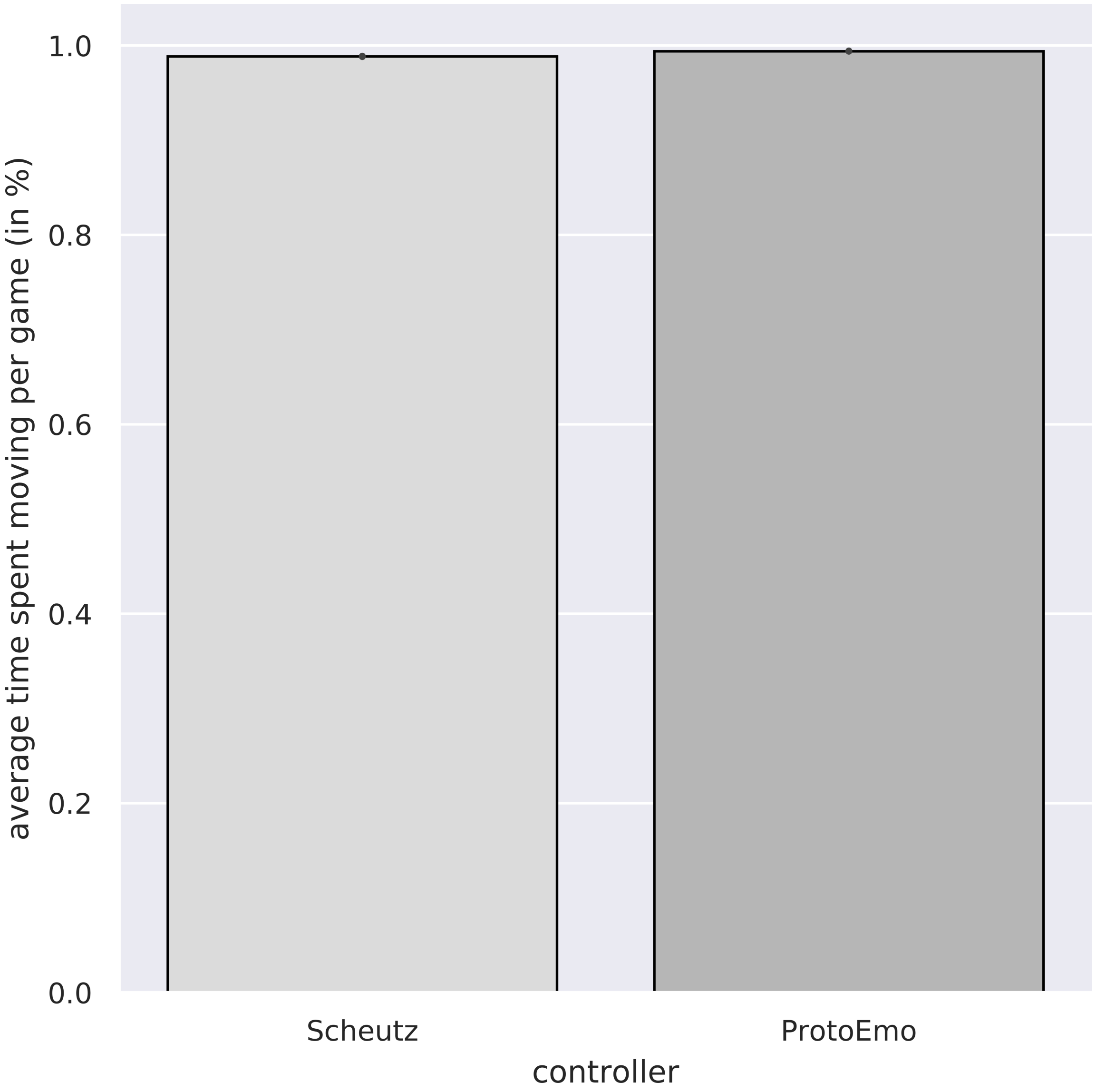


Figure 5.10 *Plot showing how much time an agent spends moving on average during a game, for each controller type. The error bar represent the* 95% *confidence interval.*

depend on them. Indeed, the different tendencies the agents use to influence the next action they will perform are computed based on the activities of the output layer. The output layer, however, is only (fully) connected to the CA layer, as shown in Figure 5.1. Therefore, even if a direct mapping between a change in tendencies and the activity of any one unit from the CA layer cannot be established, it still reflects a change in the global activation pattern of the CA layer. Consequently, to explore the veracity of the second hypothesis (H1-2, detailed in Section 5.1) Figure 5.11 and Figure 5.12 can be considered as interchangeable. Moreover, since the polynomial regressions shown in Figure 5.12 lend themselves more easily to interpretation, they will be used for the remainder of this section.

In the context of speculating about the strategies that each controller type might have employed, Figure 5.12 is especially interesting, since it highlights the individual rules each agent follows. Concerning the **ProtoEmo** controller, a few simple behaviors stand out:

- The lower an agent's energy level is the higher it forage tendency. This means that, as for animals, '*hunger*' motivates an agent to look for '*food*'.

- Supporting the results from Figure 5.6 and Figure 5.7, an agent's fighting tendency remains the same regardless of its distance to other agents. Moreover, since the value is quite low ($-10$), an agent will generally do its best to avoid any fight. In fact the only time an agent might seek out a battle with its peers is when the agent is far away from any obstacles.

- As expected, an agent's action tendency will increase the lower its energy level is. For an agent to have a high action tendency does not imply that it seeks to fight with others. The agent is simply making ready in case it finds itself in a battle situation.

- However, as the second and last diagrams on the top row indicate, the action tendency is also decreasing with the distance between the agent and the closest energy source. This supports the strategy laid out above, as it translates to the agent lowering its defenses when getting near a forage location.

- Within expectations as well, is the fact that the avoidance tendency decreases the closer an agent gets to an obstacle.

- It seems that for both obstacles and other agents, the nearer they are to the current agent, the higher the forage tendency is. It is as if agents are pulled away from dangerous situations to rather focus on looking for more energy.

Not only do the rules, outlined above, and the results from Figure 5.12 support the strategy **ProtoEmo** is speculated to have used, but they also substantiate Hypothesis 1-1 and Hypothesis 1-2. Indeed they show that, **ProtoEmo** behaves in a way similar to how basic organisms are expected to act, and that it managed to detect situations or features relevant to the agent's survival. Furthermore, with the central nucleus of the amygdala being the only layer capable of influencing an agent's tendencies, it is safe to conclude that it acts as a trigger for this survival circuit.

In the context of Hypothesis 1-3, the last set of diagrams depicting the average overall activity of the CA layer against the different input features (see Figure 5.13) are quite surprising. Indeed, from any living organism it would be expected that the closer it is from a relevant object, the higher its amygdalar activity. However, from the different plots in Figure 5.13 only the second one, describing the evolution of the average CA activity against the distance to the nearest energy source, shows an increase in activity when an agent is close to an energy source. Nevertheless, analyzing the results of Figure 5.13 within the context of the strategy outlined above, for the **ProtoEmo** controller, the last two columns do make sense. Since **ProtoEmo**'s agents have been found to avoid fighting with each other at all costs, and because inert obstacles do not represent any danger for the agent's survival, there is no need for the amygdala to be further activated the closer an agent is to either an obstacle or another agent. It is important to note that as soon as either an obstacle or another agent enters the detection range the amygdala is positively excited and remains in this state. This could be interpreted as an agent noticing a new obstacle or another agent and being aware of them at all time. Regardless of the context in which the first plot is interpreted, the data it represents goes against all expectations. Since the average activity of the CA layer is negative, when the agent's level of energy reaches zero, and is positive, when the agent's energy level is at its maximum, it means that nearing '*death*' is less important for **ProtoEmo** than being full of '*life*'. In the light of these last results, therefore, Hypothesis 1-3 has to be considered invalid.

In conclusion, using the **ProtoEmo** architecture I was able to show that the survival circuits found in animals can also be used to increase the survival potential of virtual agents. Furthermore, within the identified circuit the amygdala, especially the central nucleus, acts as a trigger detecting opportunities relevant to the agent's survival. Concerning the role of the amygdala in the '*arousal*' part of emotions it seems that further investigation is required, since the average activation value of the central nucleus of the amygdala did not have the correct correlations with the different input features. For anyone interested in reproducing the resource

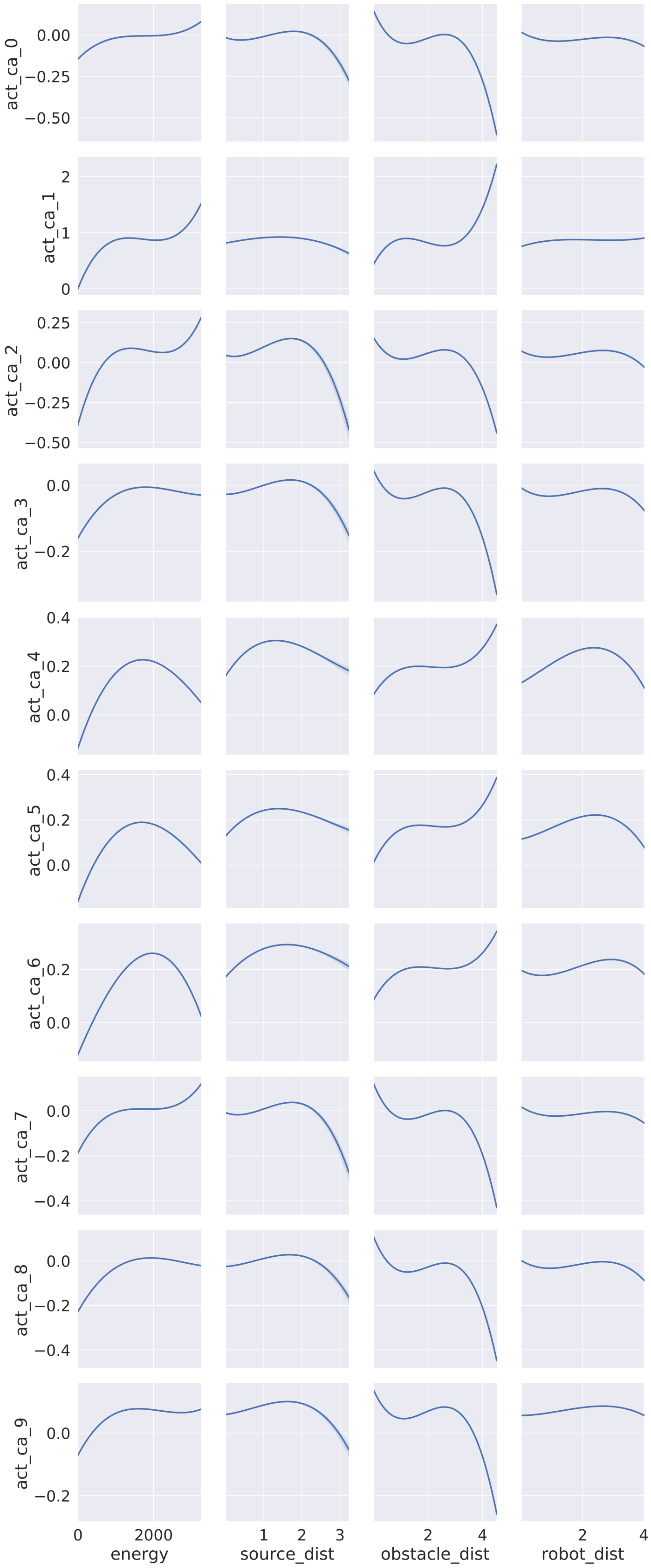


Figure 5.11 *Line plots of polynomial regressions (order=3) modeling the evolution of the activation values of each of the neural units in the central nucleus (CA) layer, against the different input features.*

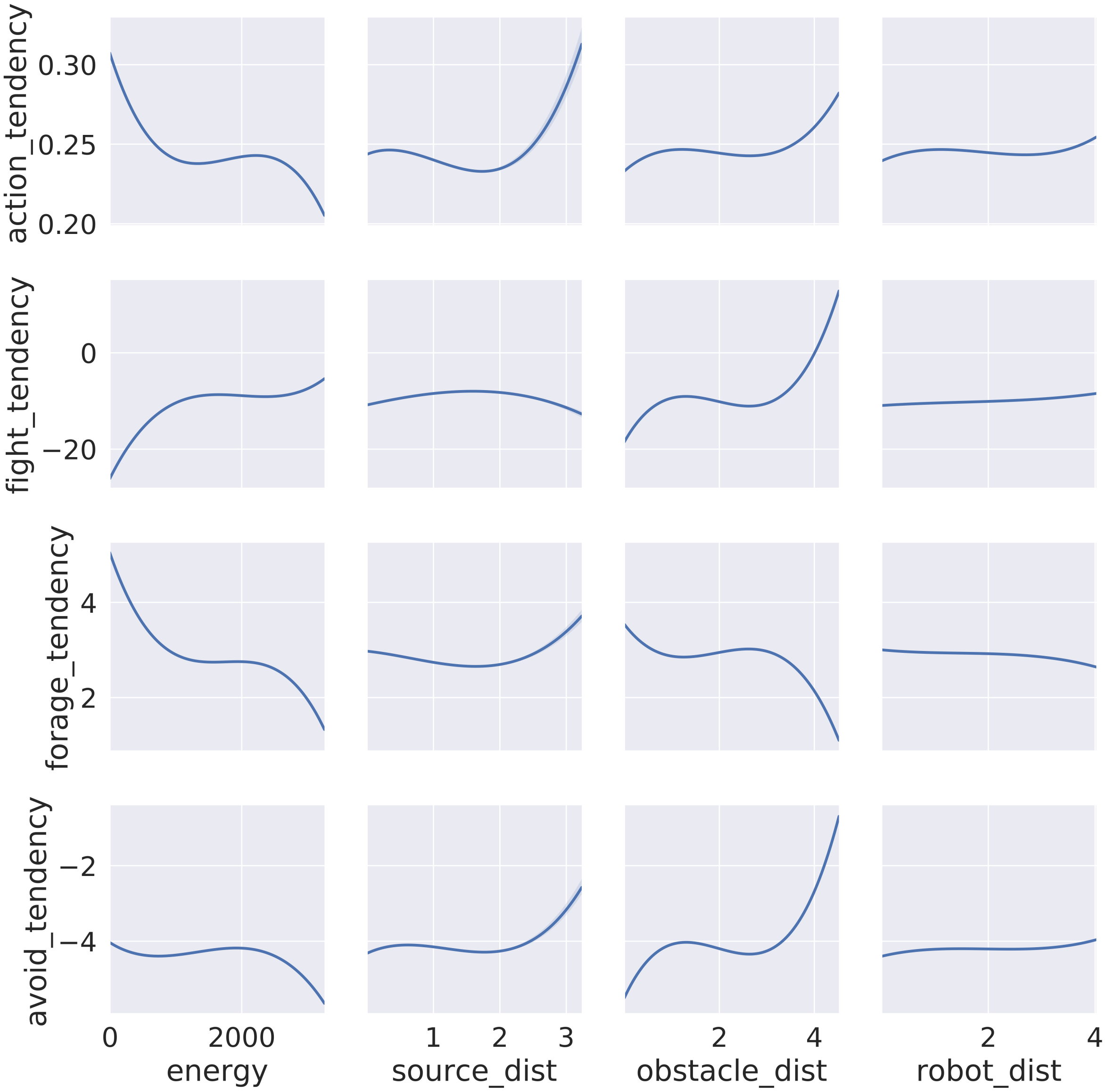


Figure 5.12 *Line plots of polynomial regressions (order=3) modeling the evolution of the average value of an agent's tendencies versus the different input features. This chart has been drawn in addition to Figure 5.11, because a change in the activation pattern of the CA layer is reflected as a change in tendencies, and tendencies are easier to interpret in terms of strategy.*

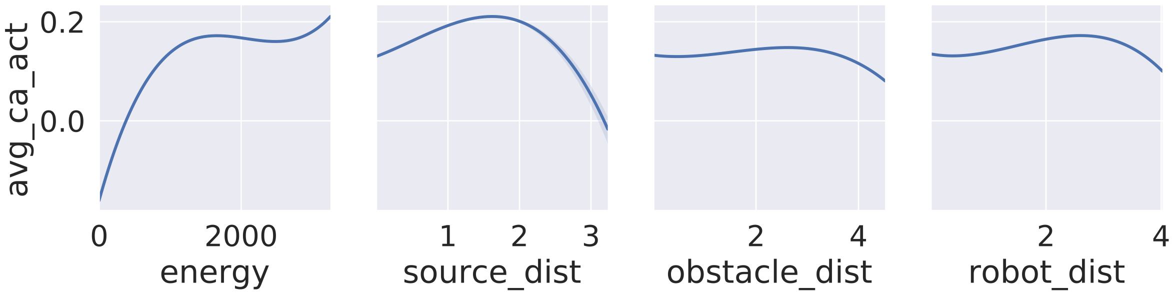


Figure 5.13 *This figure shows the line plot of the polynomial regressions (order=3), each modeling the evolution of the average total activity of the amygdala's central nucleus (on the Y-axis) against one of the inputs provided to the controllers during the resources foraging task (on the X-axis). Given that the amygdala is hypothesized to function as an alarm mechanism, detecting survival relevant situational features, interrupting ongoing processes, and reallocating cognitive resources (see H*1-2 *in the introduction to this chapter). Consequently, if the average total activity of the central nucleus increases for a specific input, it indicates that this stimulus is significant for the survival of the virtual agent.*

foraging experiment described in this chapter, the source code corresponding to the simulation, as well as both of the controllers is available on Gitlab [1].

---

[1] https://gitlab.com/davinellulinvega/epuckvsscheutz.git

# Summary

*As described in details in Chapter 2 LeDoux (2012) suggests, in his* 'Survival Circuit' *theory, that emotions are the results of the conscious mind detecting the triggering of a survival circuit. In addition, LeDoux (2012) also identified a neural structure at the base of the brain, which is essential to an animal's survival. This circuit consists of the thalamus, receiving stimuli from the environment and projecting to the lateral nucleus of the amygdala, the hypothalamus, in control of the autonomic nervous system, it sends information concerning the internal state directly to the central nucleus of the amygdala, and finally the amygdala. According to the many classical conditioning experiments involving the amygdala, it is thought that they play a major role in processing emotions. However, what type of processing and the underlying mechanisms are not well-known yet. Before being able to entertain the possibility that virtual agents can have emotions it is, therefore, necessary to first investigate if the survival circuit suggested by LeDoux (2012) can be applied to virtual agents and increase their survival potential. Consequently, I hypothesized that:*

**H**1-1 *The circuit identified as essential to an animal's survival, can also be used by virtual agents to influence their actions and increase their own survival capabilities.*

**H**1-2 *The amygdala acts as the trigger for this survival circuit.*

**H**1-3 *This survival circuit, however, is not enough to elicit any emotions. Where emotions are differentiated, from other mental states, by their* 'arousal' *and* 'valence' *(see Chapter 2).*

*To investigate the veracity of the different hypotheses, the* ***ProtoEmo*** *architecture was built and tested on a* 'resource foraging' *task. In this experiment, groups of robots using the* ***ProtoEmo*** *controller and emotional agents described by Scheutz (2004) had to survive by foraging energy sources. The robots were also capable of reproducing to increase their population, and of fighting in case two or more agents wanted to forage the same resource. However, each agent could not directly decide which action would be performed next, instead it had to rely on four tendencies to influence the decision-making process. According to the results, the* ***ProtoEmo*** *architecture was indeed able to improve the survival capabilities of the virtual agents it controlled. Furthermore, the amygdala was able to detect opportunities relevant to the agents' survival. However, it was determined that the amygdala's central nucleus could not fulfill the* 'arousal' *part of emotions.*

# CHAPTER 6

# **PrimEmo**: Primitive emotions and decision-making in the brain

**PrimEmo** (shown in Figure 6.1) is the name given to the system born from the aggregation of: **ProtoEmo**, the Primary Value Learned Value and the basal ganglia models (whose detailed descriptions can be found in Section 5.1 and Sub-Section 3.3.2 respectively). While **ProtoEmo** was inspired by the survival circuits theory, suggested by LeDoux (2012) (see Sub-Section 2.6.2), it only implements half of such mechanism. According to LeDoux:

> "*survival circuits are sensory-motor integrative devices that serve specific adaptive purposes. They are tuned to detect information relevant to particular kinds of environmental challenges and opportunities, and they use this information to control behavioral responses and internal physiological adjustment that help bring closure to the situation.*" — LeDoux (2012, p. 5)

Therefore, as hypothesized in Section 5.1, **ProtoEmo** only provides the mechanism for detecting stimuli relevant to the survival of virtual agents. If the hypothalamus were to be used as an output influencing the agent's internal state, **ProtoEmo** would also fulfill the role of '*internal physiological adjustment*' mentioned above. As a consequence, by taking further advantage of the PVLV's and basal ganglia's capabilities, the **PrimEmo** architecture fully implements LeDoux's (2012) survival circuits.

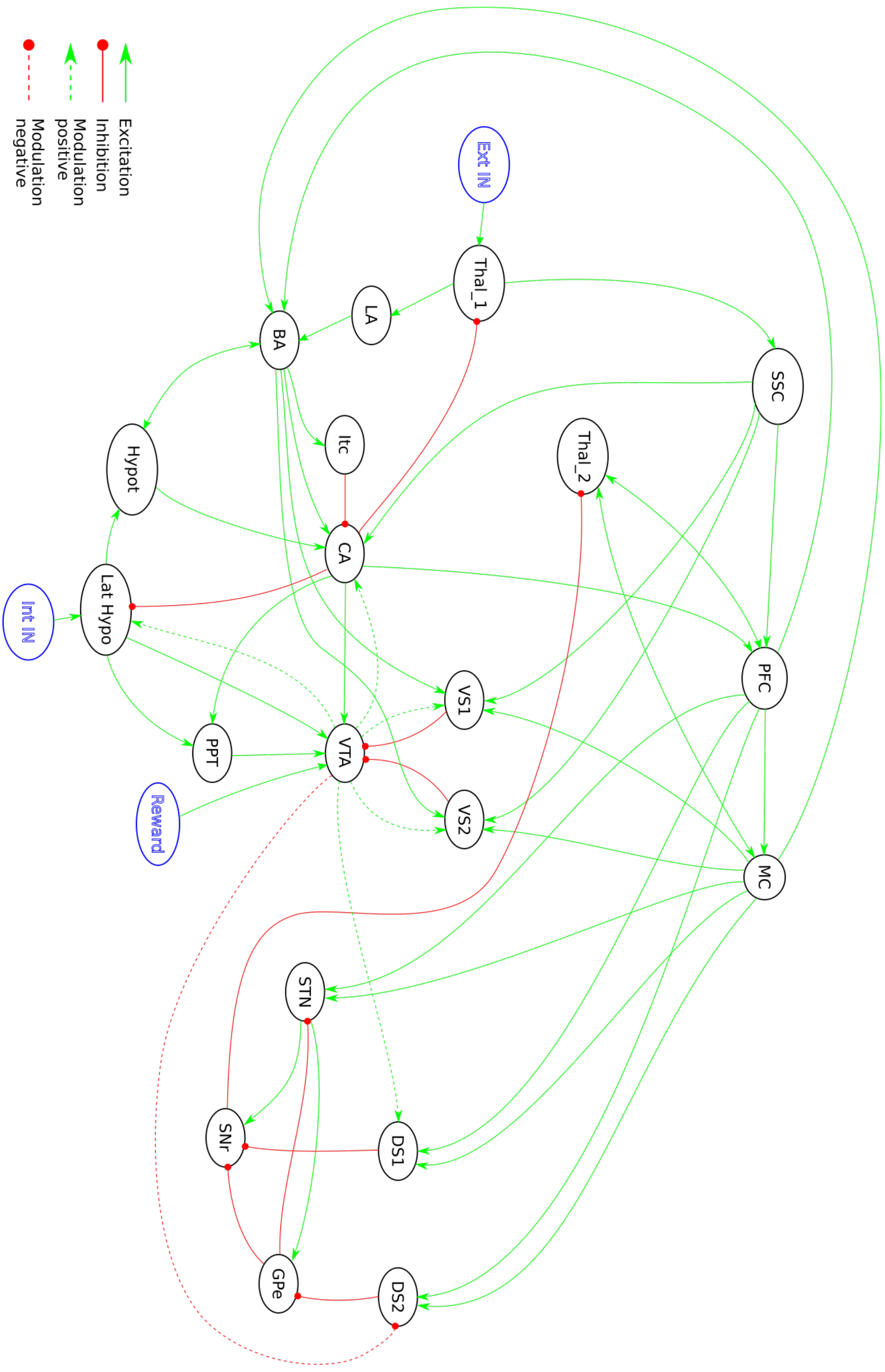


Figure 6.1 *A global representation of the* ***PrimEmo*** *architecture.*

## 6.1 The PrimEmo architecture

As depicted in Figure 6.1, the **ProtoEmo** architecture introduced in the previous chapter shares two neural populations with O'Reilly et al.'s (2007) PVLV model: the central nucleus of the amygdala and the lateral hypothalamic area. The main role of the PVLV is to activate the dopamine system for innately relevant stimuli or associated ones. Hence, it provides a reward prediction error signal to its target brain areas. In **PrimEmo**, the dopamine and serotonin systems have been merged into a single neural population, whose activation range is $[-1,1]$ instead of the standard $[0,1]$. 0 is considered to be the neutral point. Consequently, if the population's average activity (denoted $a$) is strictly positive ($a > 0$), then the PVLV's output represents a reward. Otherwise (for $a < 0$) the signal is interpreted as a punishment. As was established in Sub-Section 3.3.2, processing within the PVLV is split between the '*primary value*' and the '*learned value*' systems. O'Reilly et al. (2007) and Hazy et al. (2010) hypothesized the excitatory components of those mechanisms to be implemented by the lateral hypothalamic area and the amygdala's central nucleus, respectively. However, in their model both brain areas were represented by a single neural population. Hence, restraining the encoding capabilities of those layers. Due to its design and the fact that its inputs are spread over many processing stages, **ProtoEmo** provides a more specific detection of innately relevant stimuli, as well as those acquired through conditioning. Furthermore, the representation of the amygdala's basal and central nuclei into separate layers means that in addition to the saliences provided by the central nucleus **ProtoEmo** also sends a '*novelty*' signal to the PVLV's striatum. Thus, complementing the amygdala's main contribution to the PVLV model. Consistent with the mathematical model of the Actor-Critic, explored in Sub-Section 3.3.2, the PVLV and basal ganglia are two independent structures. Their only means of communication is through the dopamine system, whose activity is controlled by the PVLV. Using its gating mechanism the basal ganglia decides the next action to perform, based on the current state (both internal and external) and the expected reward (computed by the PVLV).

Upon receiving a punishment or reward from the environment after execution of a specific action, the basal ganglia, as well as the PVLV, use the reward prediction error signal to update the policy and value function, respectively. On the contrary to what has been the case so far, the basal ganglia system and **ProtoEmo** architecture do not have any neural population in common, within **PrimEmo**. Although, the somatosensory (SSC), pre-frontal (PFC) and motor (MC) cortices do feed back to the basal and central nuclei of the amygdala (both part of **ProtoEmo**). Following LeDoux's (1986; 1992) suggested model of the '*two pathways*' to the amygdala (explored in Section 4.1) the SSC provides more refined representations of incoming stimuli used by the central nucleus for conditioning. On the other hand, both PFC and MC project to the basal nucleus and maintain a context for the amygdala to better process the relevance of a given stimulus.

Finally, it is important to note that even though the structures making **PrimEmo** have been presented in a sequential fashion, they form a recurrent loop. Furthermore, as discussed in the previous chapter, most of the mechanisms rely themselves on feedback control systems to operate. Thus, **PrimEmo** is a recurrent architecture made of smaller control loops, and most of its activity is asynchronous. As a consequence, for **PrimEmo**'s output to be useful to an external program, it needs to be run until it reaches stability. Stability in this case is defined by the activity of the output layer being the same for more than two cycles. More details related to the exact structure and implementation of the **PrimEmo** architecture can be found in Appendix B.

Although the previous chapter set out to prove that **ProtoEmo** is quite a capable architecture in itself, it still lacks the ability to take any definitive decision. Both the amygdala and hypothalamus are limited to influencing other parts of the brain. **PrimEmo**, on the other hand, takes advantage of the basal ganglia's gating mechanism to control the '*species-specific*' reaction to an innately triggering stimulus, as mentioned by LeDoux (2012). Furthermore, through the activity of the merged dopamine and serotonin systems, controlled by the PVLV, **PrimEmo** should able to build on **ProtoEmo**'s saliences to support primitive emotions. As a complete implementation of LeDoux's (2012) survival circuits **PrimEmo**, therefore, should enable virtual agents to better survive in complex and dynamic environments, in addition to endowing them with primitive emotions.

In conclusion, the two main hypotheses explored in the remainder of this chapter, can be expressed as:

**H**2-1 The use of a system capable of encoding the salience of survival relevant situational features will enhance the survival capabilities of virtual agents in complex and dynamic environments.

**H**2-2 Furthermore, the mechanisms responsible for representing the salience of the survival relevant situational features will trigger an emotional episode. Emotions are differentiated from other mental states, by their level ‘*arousal*’ and ‘*valence*’ (see Sub-Section 2.7.3 for the exact definition of emotion adopted in this thesis).

## 6.2 Viking Doom: A competition putting PrimEmo to the test

Having outlined the **PrimEmo** architecture in the previous section, as well as dedicating a complete chapter to its different sub-components, this next section describes the experimental setup designed to observe **PrimEmo**’s performance in complex and dynamic environments. When mentioning performance, this thesis is less interested in having **PrimEmo** demonstrate its superiority, solving a given task, compared to other state-of-the-art algorithms, than it is in emphasizing the role emotions play in the decision-making process. As a result of those specific requirements, a complete game has been designed from the ground up. The need for a task (and an environment) complex enough to warrant the use of emotions in the decision-making process, is further reinforced by the fact that despite recent efforts (from companies such as OpenAI and DeepMind), there is currently a lack of standard framework for training and testing different types of agents. Here different types is taken to mean agents whose learning process is based on reinforcement learning, genetic algorithms or other neural network based techniques. The reasons for choosing a game, rather than a simulation, as a testing ground for the **PrimEmo** architecture, will become apparent in the next Sub-Section.

### 6.2.1 Why a competition rather than a simulation?

Considering that both PVLV and the basal ganglia models (described at length in Sub-Section 3.3.2) have already been proved capable of solving complex tasks (Atallah et al., 2004; Collins & Frank, 2013; Frank & Claus, 2006; Girard, Cuzin, Guillot, Gurney, & Prescott, 2003; Gurney et al., 2001b; Gurney, Prescott, Wickens, & Redgrave, 2004; Hazy et al., 2006; Hazy, Frank, & O’Reilly, 2011; Prescott, Montes González, Gurney, Humphries, & Redgrave, 2006), the ‘*one-resource foraging*’ experiment suggested in Chapter 5, would not allow **PrimEmo** to demonstrate its full potential. As a consequence, a more involved experiment, shaped as a game called: ‘*Viking Doom*’, was implemented. As the name implies, it is a viking themed roguelike game to be played by virtual agents only. Roguelike are usually characterized by a party of mortal heroes having to explore and fight their way out of a procedurally generated dungeon. On the contrary to most other game genre, in a roguelike, when the main character dies, it is not brought back to life immediately. Instead, the player has to start again from the beginning, losing all of his progress so far (a concept usually referred to as '*permadeath*'). Rooms in the dungeon are represented as a tile-based map, where player and monsters are allowed to freely roam around. Finally, the actions of both players and non-playing characters are considered sequentially, by the game engine. The idea behind Viking Doom was to organize a competition open to the public. By having **PrimEmo** take part in the competition, it would be pitched against many different strategies. With enough participants, and therefore data concerning **PrimEmo**'s behavior, Viking Doom would provide us with a solid basis to draw conclusions as to the performances of the **PrimEmo** architecture.

### 6.2.2 The game and its website

The Viking Doom project has been split in two parts. The first is the engine containing all the logic, communication protocols and objects required for playing the actual on-line game. The second, is an accompanying website, that serves two purposes. It describes the experiment's goal, and introduces the rules and the game's lore. However, its main function is to act as the display, showing what is happening within the game engine in real time (an example of a game being displayed on the website is shown in Figure 6.2 and more can be seen in Figure C.1; a live version of this display is also available in video format [1]). The whole project has been written in Python and is publicly available on Github [2]. While the website relies on the Django framework and its REST extension (Holovaty & Willson, 2017) to do the heavy lifting, the game engine at its core implements the Actor pattern (first introduced by Hewitt, Bishop, and Steiger (1973)). By definition, upon receiving a message an actor can only perform a limited number of actions: *1*) modify its internal state, *2*) send messages to other actors, *3*) and create new child actors. Since actors do not share any modifiable memory, their only way of communication is via messages, thus avoiding the need for any lock mechanism. The actor pattern is therefore ideal for implementing stateful tasks for massively concurrent systems. As the game engine can be decomposed into a set of independent tasks, that can each be assigned to a single actor, this allows it to take advantage of the parallel processing capabilities of the host server. To allow the participants to implement their virtual agents (usually called bots in this context) using the programming language of their choosing, Viking Doom uses a '*client / server*' architecture for the interaction between the game and its players. Thus, one of the actors making up the game engine, implements a simple HyperText Transfer Protocol (HTTP) server. To control their heroes, participants send requests to this server, which in turn communicates with the node in charge of updating the game's state, to finally have the corresponding hero

---

[1] `https://youtu.be/yndgfQ2HobM`

[2] `https://github.com/casparluc/VikingDoom.git`

Figure 6.2 *This figure is a simple example of a Viking Doom game being played, on the front page of its website. In the middle of the viking themed frame, is the map with the four heroes, some enemies, mines, markets and items. All the movable elements, such as the heroes, skeleton units and upgrade caravans, are updated in real time, thanks to the Websocket client handling the bi-directional communication in the background. At the bottom of the frame, the statistics, relevant for each of the four heroes, are displayed. The top line is the amount of gold gathered so far, while the bottom row indicates the level of health.*

move. As an answer to their requests, players receive the new state of the game as a JavaScript Object Notation (JSON) formatted object. As a consequence, the implemented bots are simple clients deciding on the next move based on the data received from the game's server. Due to the client / server nature of the architecture and the Actor pattern at the core of the engine, players can move their heroes asynchronously from each other. Consequently, neither players nor enemies have to wait for others to act before being allowed to do so themselves. Since the game itself is hosted on a remote server and manages requests from connected clients, it has to be able to gracefully handle players disconnecting at any point in time. This is the main reason behind the game having a set pace (the game completes a cycle every second) and players' requests being processed asynchronously. However, it also has important consequences for the bots' design. To win a player has to implement a strategy that strikes a balance between being complex and it being fast. Finally, in the case of a competition it gives an advantage to participants with fast Internet connections and a lot of computing power. Nevertheless, those advantages are negligible when compared to the ones afforded by a better strategy.

A typical game is characterized by a maximum number of 1000 cycles, a set of four players, each controlling its respective hero and a map. Obstacles on the map are static and have been drawn by hand on the background. To avoid any bot design taking advantage of the map's layout, twelve maps have been configured and one out of those is chosen at random at the beginning of each game. To make things more challenging for the players, upon initialization a map is populated with:

- Four heroes, one for each of the players taking part in the game. Heroes are characterized by a level of health, a level of strength and a level of gold. The goal for all heroes is to gather as much gold as possible, which can be done by either collecting purse and chest items or by defeating orcs and gaining ownership of many mines. Engaging in combat has the potential for increasing the hero's strength level, as a reward for defeating its opponent, but comes with the cost of losing health in case the hero takes damage. A simpler, but less cost effective solution for increasing a hero's strength and regaining some life is to buy upgrades and potions from the different markets. Upon death (defined by a level of health equal to zero) a hero loses ownership of his mines and is reincarnated at his original position. His level of gold and strength, however, remain as they were before his demise.

- Different enemy units, of which there are three types: skeleton, orc and dragon. The sprites corresponding to the enemy units can be found on the second row in Figure 6.3.

Of all the enemy units, skeletons are the weakest, but they are also the only mobile units. At the end of each turn of the game, all skeletons move in a random direction. Due to its overwhelming strength there is only one dragon unit per game, which stay in the same location at all time, patiently waiting for heroes to attack. In terms of enemy units, the orcs are a bit different in that each unit is tied to a mine. The role of the orcs is to guard the mines, making it difficult for heroes to acquire them. Regardless of the type, by defeating an enemy a hero is always reward with both strength and gold units. Furthermore, dead enemies disappear from the map and are automatically replaced at the end of each game turn. Each orc being tied to a specific mine, once defeated it will reappear to guard the interest of its conqueror. Skeletons and dragons, on the other hand, are scattered randomly around the map.

- An assortment of items to be collected. The type of a specific item is chosen among the following possibilities: purse, chest and health potion. All three types of items are depicted on the right of the hero, on the bottom row of Figure 6.3. Upon collection, both purse and chest items increase the gold level of the hero by a set amount, while potions heal the hero's wounds, if any. Collected items are automatically replaced by other items at the end of each game turn. The locations and types of the new items are chosen randomly.
- Two types of markets: upgrade caravans, which allow the hero to gain strength in exchange for some gold, and potion markets, from which health can be purchased. An example of the sprites for both the market and caravan, as well as the mine, described below, are shown on the top row of Figure 6.3.
- And finally mines, which can be acquired by defeating the orc guarding them. Once owned, a mine contributes five units of gold to the owner on each turn of the game.

Even though, the game engine is really the heart of the Viking Doom project, what matters more for the user is the second part, namely the accompanying website. Although, it is made of only five pages, the website contains everything a participant would need to understand the rules of the game and implement his own bot. The home page, welcomed any visitor with a screen (implemented in JavaScript) displaying one of the games currently taking place (see Figure 6.2 and Figure C.1 for examples of this display). It also allowed further navigation to the project's back story (adding a fantasy element to the whole game) and rules, as well as a detailed technical documentation explaining the communication protocols and data to be expected when writing a controller. An on-line form was also included so that new participants

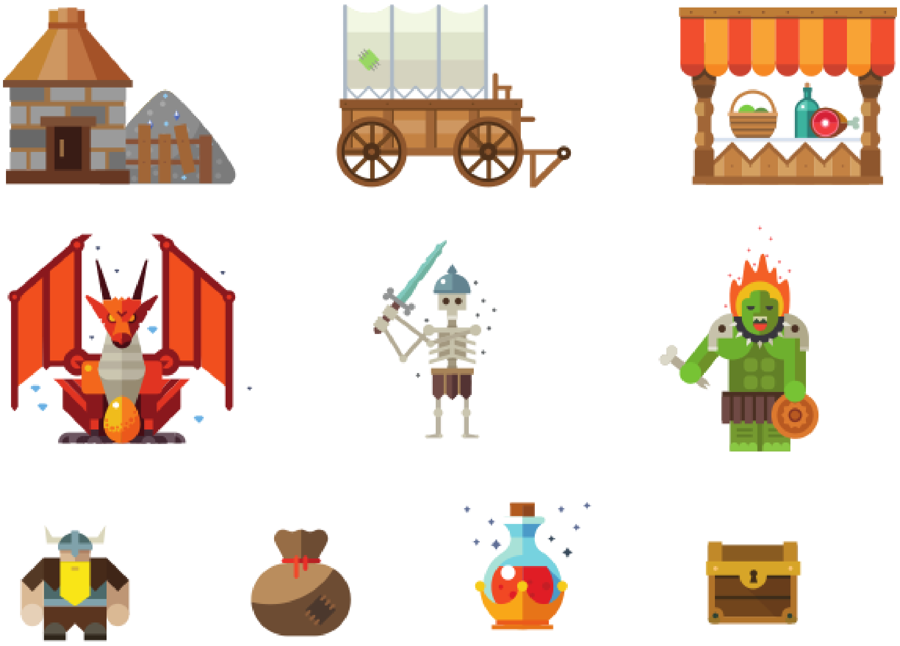


Figure 6.3 *These are the main sprites used in the Viking Doom game. From left to right and top to bottom they are: the mine guarded by the orc, which provide gold to its owner; the upgrade caravan, providing strength upgrade to the heroes; and the potion market, where heroes can replenish their health. On the next line, the different enemies are: the dragon, which is the most powerful enemy, skeletons, the weakest and most plentiful and finally orcs, that protect the mines. Finally, the last line contains the different items and an example of a hero. The purse and chest items both contain gold, but in different amount, while the potion grants some health back.*

could register as players for the games to come. Each player was attributed a single identification number upon registration. This helps the game engine to keep track of the relations between participants and heroes. More importantly, thought, this same number is also used to link a player to his overall rank. To make the competition more interesting and motivate visitors to join, a leader board is included below the screen on the home page. Finally, the last page on this domain explained the reasons behind the Viking Doom project and the goal of the PhD project it is a part of. More details concerning the interface developed for players to implement their own virtual agents, as well as the protocol adopted for communicating between the players and the game engine are presented in Section C.1. Additionally, if you wish to run your own version of Viking Doom the source code is available for download on GitHub [3].

Despite all the advertisement and the posts on different social platforms, related to AI or Machine Learning (ML), the competition was not as successful as expected and did not yield any useful results. Unfortunately, only a few people took part in the month long competition, but even those were only here to play around and did not make a real attempt at competing with each other. Still an effort was made to try and still have the **PrimEmo** architecture compete against a standard Q-learning algorithm and an Actor-Critic agent. However, it quickly became apparent that the game setup itself was fatally flawed. Indeed, since the different items were appearing at random location on the map and a hero only had to stand on a neighboring tile to collect an item, it means that a valid solution to the problem of gathering gold while staying alive, is for a hero to simply stand in place, doing nothing until the end of the game. While it is true that this is a valid solution, it is far from optimal. Furthermore, because this thesis is more interested in the agent's strategy than it is in the final level of gold gathered, needless to say that a motionless hero is not ideal for observing the role of emotions in decision-making. There is one positive point though. The fact that all three agents, at some point or another, converged toward this specific solution is interesting in its own right. At any rate, having only a limited amount of time to complete this experiment, the issue was dropped and the first version of the Viking Doom project with it.

---

[3] `https://github.com/casparluc/VikingDoom.git`

## 6.3 Viking Doom version 2.0: Experimenting with PrimEmo alone

Given the limited time frame and available resources implementing a great number of unique strategies, to push the **PrimEmo** architecture to its limits, was not realistically feasible. Therefore, it was decided to rebuild the game engine from the ground up (fixing the previous design mistakes along the way) and compare the architecture to only two other types of agents: a Deep Q-learning and an Advantage Actor-Critic (A2C). At the time, those were considered state of the art in reinforcement learning, but have since been bested by other methods, such as Proximal Policy Optimization (Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017) or Deep Deterministic Gradient Descent (Lillicrap et al., 2016). Regardless, the remainder of this section describes the new game engine, its rules and the different objects populating the map.

### 6.3.1 A new game engine, a new set of rules

Similar to the first version, Viking Doom is still a viking themed roguelike game, in which heroes have to gather as much gold as possible, while surviving in a complex and dynamic environment. Unlike its predecessor,however, this version is much closer to a true roguelike. Retaining the tile-based map and dungeon crawling elements introduced in the first version, this new game engine drops the asynchronous communication (and by extent the asynchronous movements of the heroes) to implement a turn based system and permadeath. As explained in Sub-Section 6.2.1, permadeath is a video game term describing the fact, that whenever the main character dies the game is over. Depending on the game, all progress made can be lost and the player usually has to restart from the beginning. The new game engine adheres to this idea, which means that when a hero's health points go down to zero, the hero dies and does not come back to life. As a consequence, there are two conditions in which the game ends: *1*) Either all heroes are dead, *2*) or the game has reached 1000 turns. This condition has been set to prevent games from running indefinitely. For a game a turn is defined by the execution of the following sequence of events:

1. Observe the environment's state and provide the features describing this state to all heroes taking part in the game.
2. Each player then decides what action his respective hero should perform next, based on the provided features. The players' decisions are sent to the engine. Once all action requests have been gathered, the engine considers them in random order. This is to avoid any player taking advantage of the sequential nature of the main loop.

3. A player can only choose an action from among the following five: Up, Down, Left, Right and Idle. If he chooses to move in one of the four directions, the game engine tries to move the hero. Invalid moves (moving on an obstacles or out of the map) are still considered as action. After each movement and before ending its turn, a hero collects any item present on the tile. In case a player decides not to move his hero (done by requesting Idle), the game engine will automatically try to make it fight against either the closest hero or the closest enemy unit, if any. Otherwise, the hero does nothing and the turn ends.

4. To allow heroes to run away from a fight their actions are considered first. After all heroes are done moving around, fighting or idling, it is time for any enemy units that can act to do so. As will be detailed later, there are two types of enemy units that can move around to target heroes passing by: skeletons and dragons. The actions of enemy units are considered in a similar fashion, movement first, battle next. For static enemies, such as Trolls, the movement part is entirely skipped, hence living them with only two possibilities: fighting or doing nothing. Dead enemy units are removed from the environment on this step as well.

5. The whole environment having moved a step forward, new enemy units and items are scattered around the map if necessary, before ending this game's turn and moving on to the next.

Even though, the asynchronous nature of the game engine was restricted via the implementation of a main sequential loop, the architecture at its core still relies on multiprocessing to take full advantage of the processing capabilities of the CPU executing the program and overall accelerate the experiment. In fact both the players, controlling the heroes, as well as the enemy units are executed in child processes, which communicate with the game engine via a set of queues. This not only facilitate the development of new controllers for the heroes, it also allows for them to learn and decide of the next action, simultaneously. The same is true for the different mobile enemy units. Parallelizing any independent task means that the game engine spends less time waiting for requests and more time actually moving the game forward.

### 6.3.2 Designing a complex and dynamic environment

The description of the main loop in the previous Sub-Section, makes it look like the game engine is doing all the heavy lifting, when in fact it is the environment's responsibility to manage all the objects populating it. While the game engine orchestrate the communication between the different processes and manages the controllers (linked to the different heroes), the environment handles the actual game. The environment is in the ideal position to perform this task, since it contains and has access to all the objects that appear on the map.

As backdrop for the adventure happening within the environment, a tiled map of size 10 by 10 is used. On the contrary to the first version of the Viking Doom project, obstacles are not hand drawn on the map, they are defined in a configuration file and later rendered over the map. Although the first steps of the experiment use a smaller map, the full game is intended to be played on a 50 by 50 grid. To simplify the rendering process the different types of obstacles (tree, water, house, lava, rock and grass) present in the previous version have been abandoned in favor of a single type, represented by a stylized rock tile. Furthermore, the edges of the map are also considered as obstacles or rather tiles that should not be stepped on. This will become important later on, when describing the formula used for computing the reward for an agent's action. Sub-Section C.2.1 shows an example of configuration file, as well as explains how each parameter influences the map's design.

### 6.3.3 Danger and incentives

To make the game more complex, dynamic and, therefore, more interesting in terms of strategy, incentives and dangers have been introduced in the form of items to collect and enemy units to avoid or fight. To avoid any strategy where the hero can '*win*' the game without having to move, to collect an item or fight an enemy the hero has to stand on the same tile as the other object. Furthermore, items and enemies cannot be initialized on an already occupied tile.

Only two types of items can be found on the map. Gold items, such as purse or chest, when collected increase the hero's level of gold. This item type has been introduced to create a dense virtual reward from which the different reinforcement learning based controllers can efficiently learn.

The second item type is the health potion. As the name implies, collecting this kind of item increases the hero's health. This item is a required resource for the hero's survival. Since fighting implies taking damage, potions are the only way a hero can get back a portion of its health and avoid death. Collecting those items then is a real incentive, compared to collecting gold ones.

In the viking doom project, danger comes in the shape of different enemy units. As is the case for the items, there are three different types of enemies, characterized by their level of health and strength. The weakest of them all are the skeletons. Skeletons are intended to be the entry-level enemy units that heroes meet early on in the game. Since they are quite weak, it allows the controllers to experiment with different strategies to deal with enemies. However, to avoid repeating the same mistake as in the first version of this project (where a player could gather gold and fight enemies by simply staying still for the whole game), the behavior of skeleton units are not random anymore. Instead, each unit is assigned an initial position and an area of effect. Whenever a hero is within the area of effect, it becomes a target and the skeleton moves toward the hero's position, looking for a fight. If the hero dies, or exit the area of effect, the skeleton goes back to its initial position and waits for another hero to come by.

The next type of enemy is the troll. As the orcs in the previous version, trolls are found close to the mines that they guard. In this version, however, mines have no purpose and are only present for the sake of presentation. Trolls are further characterized by the fact that they are static. This means that for a hero to fight with a troll, it has to voluntarily step on the tile the troll is already occupying. The main reason for introducing this kind of enemy unit, which is not all that threatening, is to create virtual obstacles or bottlenecks in some areas of the map.

The final and strongest opponent a hero can encounter on the map, are dragons. Dragons are quite the fierce opponents capable of ending the life of any hero in only four turns. Dragons have the exact same behavior as skeleton units. As a result of their massive built, however, dragons have slower movements. This means that they can only move every two turns of the game. Still, they can fight on every game turn.

During a fight, the amount of damage dealt is equal to the enemy's strength level and is subtracted from the hero's current health level. Furthermore, upon death the enemy unit is deleted from the map, by the environment, and replaced on the next game's turn. The different sprites used for representing the enemy units, as well as the heroes and items remain the same as the ones shown on the two bottom rows of Figure 6.3. Additional details required to implement the full game engine are described in Sub-Section C.2.2.

### 6.3.4 Heroes and controllers

With the environment firmly in place, it is time to focus on the main character of this game, namely the hero and its associated controller. As mentioned in Sub-Section 6.3.1, to minimize the time required to go through a single turn of the game, the engine takes advantage of the multiprocessing capabilities afforded by the Python language and the multi-core architecture of the host's CPU. As such any task that could be independently run has been delegated to a child process, managed by the engine. Controllers are such tasks. Indeed, once the environment has moved one step forward, each controller can independently observe the new input state (made of both the environment's and corresponding hero's states), learn and decide on the next action to perform. Rather than having controllers wait for each others' decisions and waist precious compute time, it is more efficient to implement them in different processes. Upon initialization each controller is attributed: *1*) a hero, which in a sense embodies the controller, *2*) and two queues, used for communicating with the game engine. One for receiving the environment's state on each turn and the other to send the next action to the game engine.

The last entity appearing on the map, but not described so far is the hero. A hero is a simple container, managed by the environment and read by its associated controller. It is characterized by a level of health, a level of strength and a level of gold gathered so far. The environment, however, also tracks the number of items collected for each item type, as well as the number of enemies fought, for each enemy type. To easily keep track of those numbers for each hero on the map, the different counters are all stored within the hero itself. Upon initialization, a hero has 100 units of health, a single unit of strength and no gold. A hero can only move Up, Down, Left or Right one tile at a time, or remain Idle for a turn. In case the controller decides that its hero should remain Idle, the game engine will automatically check for the possibility of having the hero fight against either an enemy or another hero. If none of the above can be applied, then the hero will truly stand still for a turn. However, to avoid any problems with heroes idling for too long, the reward provided to the controller's learning process, includes a $-0.001$ term. This term is a sort of incentive for the hero to always move and at the very least explore its environment. For animals, this could be interpreted as the cost in energy to maintain the body's homeostasis.

In addition to being implemented as processes, all controllers within the Viking Doom project are required to inherit from a generic interface. A simple and common Application Programming Interface (API) has been designed to ensure that any controller, that wishes to play the game, has access to exactly the same information (concerning the state of both the environment and associated hero). In the context of this thesis, a common API means that all experimental parameters are the same for all controllers, and only the learning and decision processes can affect the outcome of the game. In addition to ensuring that all things are equal for everyone, this generic interface also implements a standard loop for the hero. The generic controller loop, repeats the following steps until either the end of the game or the death of the associated hero:

1. Based on the current or initial state of the environment and that of the associated hero, decide the next action to perform. The chosen action is then sent to the game engine.
2. Next the controller waits for the environment to move forward. Once this is done, the game engine broadcasts the environment object, containing the new state, to all controllers.
3. From this environment object the controller extracts the input features, which are then used by both the learning and decision-making processes.
4. Based on the same environment, the controller also computes the reward associated with the action chosen on the first step.
5. Finally, using the different extracted input features and the computed reward, the controller learns from its mistakes.
6. To be able to extract any strategy or observe the role of emotions in the decision-making process, an additional step has been introduced. This last step monitors the state of both the environment and the hero. All monitored data are written to an external file for further processing to extract the results discussed in the next section.

Even though a standard control loop has been implemented for all child classes, a specific controller can still overwrite this loop to fulfill more specific needs.
As mentioned multiple times already, the decision-making and learning process rely on the environment's state and the levels of the hero's attributes. More precisely, the input features extracted by the controller on each loop are:

- Enemies: For the two nearest enemies, the controller is told about their positions, as well as their respective health and strength levels. Knowing the enemies' positions is fundamental for the controller to be able to avoid such danger. Rather than dedicated a single flag to determine the type of enemy, providing both the health and strength means that each hero can decide to either fight or flee, depending on its own internal state.

Consequently, a hero at full health and strength could very well decide, for example, to try defeating a dragon already been weakened by someone else.

- Items: The positions of the two nearest gold items and health items are provided to the controller as part of the environment's state. For the gold items, however, the controller is not told of their values (a purse is worth 5 gold and a chest 25). This is so that, when making a decision the agent cannot expect a set reward, and each choice has to be made based on past experience and a careful evaluation of the situation.
- Heroes: If any heroes are present, other than the one managed by the current controller, the positions, health, strength and gold levels of the two nearest are extracted from the environment's state. Similar to the enemies, a controller has to know the exact location of any potential dangers to be able to avoid them. Although heroes are not separated in to different types, their levels of strength and health can evolve during a game. Therefore, for a controller to decide whether it should fight or avoid a given hero, it has to know if it can be defeated (this depends on its strength and health) and if this is worth the risk (this is indicated by its gold level).
- Obstacles: To avoid the current hero getting trapped in an inescapable situation, the positions of the three nearest obstacles are given to the controller.
- Current Hero: Finally, the hero's internal state is represented by its levels of strength, health and gold. The hero's current position is not required and might lead to controllers learning the exact position of every event, rather than generalizing from experience.

It should be noted that the different positions described above are all relative to the hero's current position. This is to avoid having to provide the current hero's position in addition to all the other inputs and hopefully make it easier for the controllers down the line, since none of them will have to learn how to compute the relative location of a given object. Moreover, to avoid any gradient explosion, when using the back-propagation algorithm for learning, all features have been normalized to the range $[0, 1]$.

The last element, that the generic controller computes and provides to the specific controllers (detailed in the next Sub-Section), is the reward. For each action taken by the hero, a corresponding reward is computed based on the new state of both the environment and the hero. The general idea behind the formula used for computing the reward is primarily to guide the controllers toward the game's main goal: collecting as much gold as possible, while staying for as long as possible. However, along the way some hints are also given to the controllers to facilitate the learning process. That is the main reason behind the inclusion of both the health and strength levels in the reward. Strength is used to signal the controller that enemies can be defeated and the more you do the stronger the enemies you can take on. Health, on the other hand, acts as both a punishment and a reward. Heroes are rewarded for collecting potions, which increase their health, but punished for fighting against enemies and with each others (an activity that decreases a hero's level of health). Health, more than gold and strength, teaches the controllers about the different dangers and how to survive in this complex and dynamic environment. The final formula used for computing the reward normalizes its value to the range $[-2,2]$ to avoid any gradient explosion, as follow:

$$reward = \frac{gold_{curr} - gold_{prev}}{25} + \frac{hp_{curr} - hp_{prev}}{25} + \frac{strg_{curr} - strg_{prev}}{10} - 0.001 \tag{6.1}$$

Where $X_{curr}$ and $X_{prev}$ correspond to the variable's current $(t)$ and previous $(t-1)$ values, respectively. Furthermore, as mentioned above a $-0.001$ term has been added to the reward to force controllers to optimize their paths between rewards and avoid any strategies, where the hero survives by simply staying idle for the whole game.

Given the limited resources available and time frame for running the experiment, comparing **PrimEmo** to a great number of unique strategies was not feasible anymore. Therefore, it was decided to compare the architecture to only two other types of virtual agents. The first one controlled by a Deep Q-learning algorithm, while the second one implements an Advantage Actor-Critic (A2C). The specifics required for the exact implementation of those controllers are presented in Sub-Section C.2.3.

For the **PrimEmo** architecture, the learning process is not as straight forward as the other standard models. This is mostly due to the fact that within **PrimEmo** there are many recurrent connections, as depicted in Figure 6.1. Although, a version of the backpropagation algorithm exists, which is compatible with recurrent units, such as Long Short-Term Memory (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Unit (Cho et al., 2014), it requires each recurrent loop to stop after a given number of steps and for the user to keep track of the network's state at each time steps to be able to compute the error. However, in **PrimEmo**'s case the controller waits until the output activity is stable, before making a decision. Here, the stability of the output layer is defined as: each neuron having the same activation value (down to a certain precision) on two following cycles. Since it is impossible to predict how many time steps it will take to reach such equilibrium, learning via backpropagation is not feasible.

Furthermore, keeping track of all the network's states would require a huge amount of memory and processing power, thus rendering the whole process ineffective and quite costly. As an alternative to backpropagation, a simple Genetic Algorithm (GA) has been implemented. As a metaheuristic GAs have some advantages over the standard backpropagation:

1. While backpropagation based methods, can easily get trapped within local optimums, thanks to the way they explore the space of possible solutions GAs are capable of back tracking to find a path to the global optimum. However, GAs are very slow to converge to said optimum, if they converge at all.

2. Since GAs use a simple fitness term, attached to each individual a given population, for their selection process, it means that individuals can be evaluated independently from each other. This in turn means that GAs are massively parallelizable and, therefore, cost effective, when run on high performance computers.

3. From a more subjective point of view, GAs are easier to implement and debug. This, though, is becoming decreasingly relevant, because of the increasing availability of libraries such as PyTorch and Tensorflow, which handle the heavy lifting.

4. Finally, even though some scientists have tried to map backpropagation to learning processes within the brain, it is still a divisive topic within the scientific community. Thus, GAs, which are actually inspired by the theory of evolution, are more biologically plausible. Moreover, using GAs for learning within the **PrimEmo** architecture draws a nice parallel between animals evolving emotions to survive and now virtual agents evolving emotions of their own.

Due to its ease of use and its efficiency in scaling computation across multiple nodes, the DEAP library introduced in Sub-Section 5.2.3 was used once again to implement **PrimEmo**'s learning process. Finally, considering that all three architectures (DQN, A2C and **PrimEmo**) have been built in populations of neural units and to simplify the implementation process all controllers use the Keras framework (Chollet, 2015), backed by the Tensorflow library (Abadi et al., 2015), for the different neural layers. More details on the exact implementation of both the **PrimEmo** controller, as well as the Genetic Algorithm used for the learning process are available in Appendix B. Furthermore, the source code corresponding to the Viking Doom project, as well as all the controllers described so far, can be found on Gitlab [4].

---

[4] `https://gitlab.com/davinellulinvega/vikingdoomv2.git`

### 6.3.5 A display to monitor the learning process

A problem that arose quite late in the first Viking Doom experiment was due to the controllers (all three of them) exploiting the randomness built into the first version of the environment. As described in Section 6.2, the goal for all controllers was to find the optimal strategy allowing them to collect as much gold as possible, while staying out of harms way (survival was not an issue at the time). However, because the different mines, a player could own, were providing a constant source of reward, and because the movements of the different enemy units were truly random, as opposed to limited to a given area, it created an unfortunate chain of circumstances, where a locally optimal strategy was to just stay put and wait for gold chests and purses to appear around the player. Since the number of gold items within the environment was higher than the number of mobile enemy units, the probability of getting attacked, while standing on a single tile, was quite low compared to having an item appear in one of the four neighboring tiles. In and out of itself, it is quite interesting to see that all three controllers managed to learn a similar sub-optimal strategy, given the same environment and set of rules. For the purpose of this research project, though, it makes for very poor results and needed to be addressed.

Two solutions have been put forth to remedy this issue. The first one was to rethink Viking Doom's design from the ground up. This resulted in the game described in Section 6.3, in which mobile enemy units are tied to a certain area of attack and gold items cannot appear on the same tile as a hero. As a result, if the hero does not move, its environment remains unchanged and no reward is received. In fact, a $-0.001$ term has even been added to the reward formula (see Equation 6.1) to prevent this scenario from happening again. The second solution was to implement a display (an example of which is shown in Figure 6.4), so that the user can visually monitor what the controller is currently doing. The rationale behind this display is, that for a human it should be quite easy to notice if a controller is developing some peculiar strategy, like not moving at all. As can be seen in Figure 6.4, the display is very simple. Its goal is only to show the latest state of the environment, not to interact with the experiment. Considering that the experiment is going to be run on a high performance computer (HPC), the display is build on top of a client server architecture. The server part is living, on the HPC, right next to the game engine, which feeds the server up to date information concerning the game's state. The server is then in charge of simply sending this data to all connected clients. Between the server and its clients bidirectional communication relies on SocketIO, a Python websocket framework. Thanks to the capabilities afforded by SocketIO, clients can connect to a server from anywhere in the word, as long as there is an Internet connection. No login is required. On the client side, the actual display is managed

Figure 6.4 *An example of the display built for monitoring a game of Viking Doom. A live version of this display can also be seen at the following url:* `https://www.youtube.com/channel/UCXALLsPnKf1cNJ0opIPAKNw`

Table 6.1 In this table are detailed the different statistics corresponding to the average life span of each controller.

| | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ctrl | | | | | |
| A2C | 325.212 | 400 | 326.366 | 357.196 | 293.229 |
| DQN | 718.500 | 400 | 394.887 | 757.199 | 679.801 |
| PrimEmo | 327.935 | 400 | 349.940 | 362.229 | 293.641 |
| Random | 139.007 | 400 | 142.326 | 152.955 | 125.060 |

via Pyglet. Pyglet is a windowing and multimedia library, which allows to easily instantiate a window and build within it complex animations. Owing to this display, one can easily monitor what is currently happening within a game, whether it is running locally or on a remote HPC.

## 6.4 Results

When considering the performances of the **PrimEmo** controller compared to the other two competitors, it is expected to have similar results to the ones using an Advantage Actor-Critic, but better than those controlled by the Deep Q-learning algorithm. The logic behind this statement stems from the fact that by definition the DQN learning process only updates its value function (as explained in Section 3.2) and has a fixed policy. On the contrary, both **PrimEmo** and the A2C simultaneously optimize their value function and policy (as detailed in Section 3.3 and 6.1). They are, therefore, able to change their strategies on the fly and possibly take advantage of environmental features to build a better solution to the task. Since the **PrimEmo** architecture and A2C model operate on the same principles, they should perform similarly. For this particular implementation (detailed in Appendix B), **PrimEmo** may have a slight edge over the agents using the A2C model, because of the difference in depth. Indeed, for this experiment, the A2C has been implemented in a network made of six layers, whereas **PrimEmo** has 20. Thus, in theory, **PrimEmo** is capable of extracting more features or concepts from the same sensory inputs, than the A2C. Consequently, with more detailed information to rely on **PrimEmo**, should devise a different and possibly better strategy.

Looking at Figure 6.5 and Table 6.1, it seems that both the prediction above and Hypothesis 2-1 might be entirely misled. Indeed, in Figure 6.5, as well as in Table 6.1 the DQN controller clearly has better survival capabilities than both the A2C and **PrimEmo**, which are performing

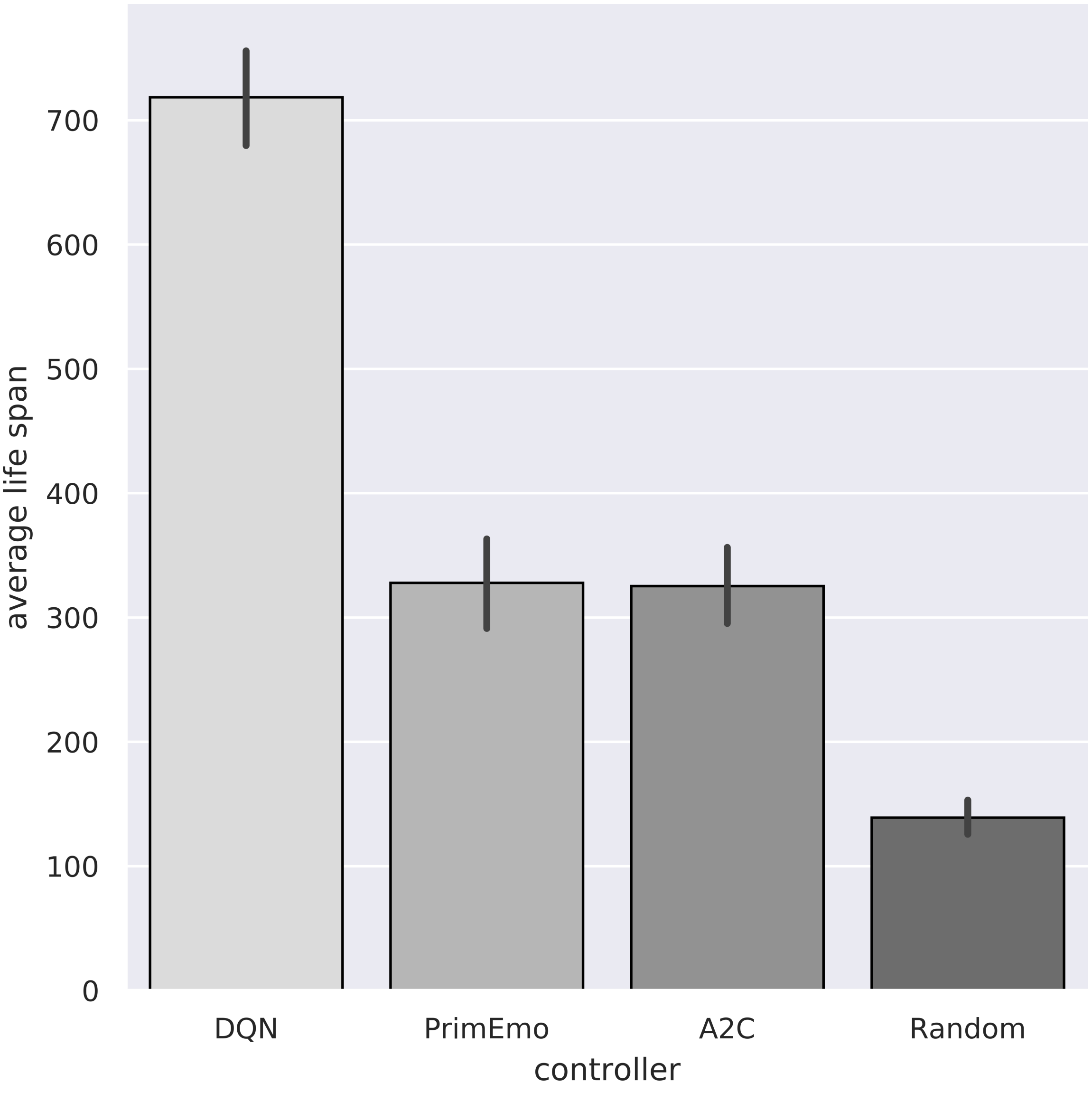
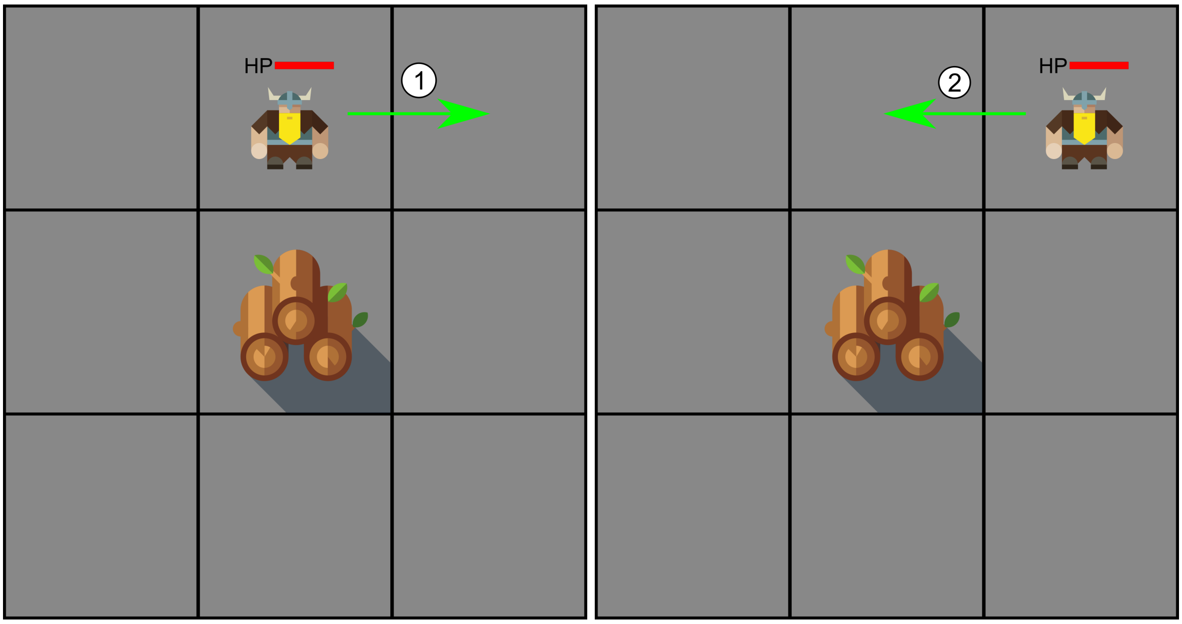
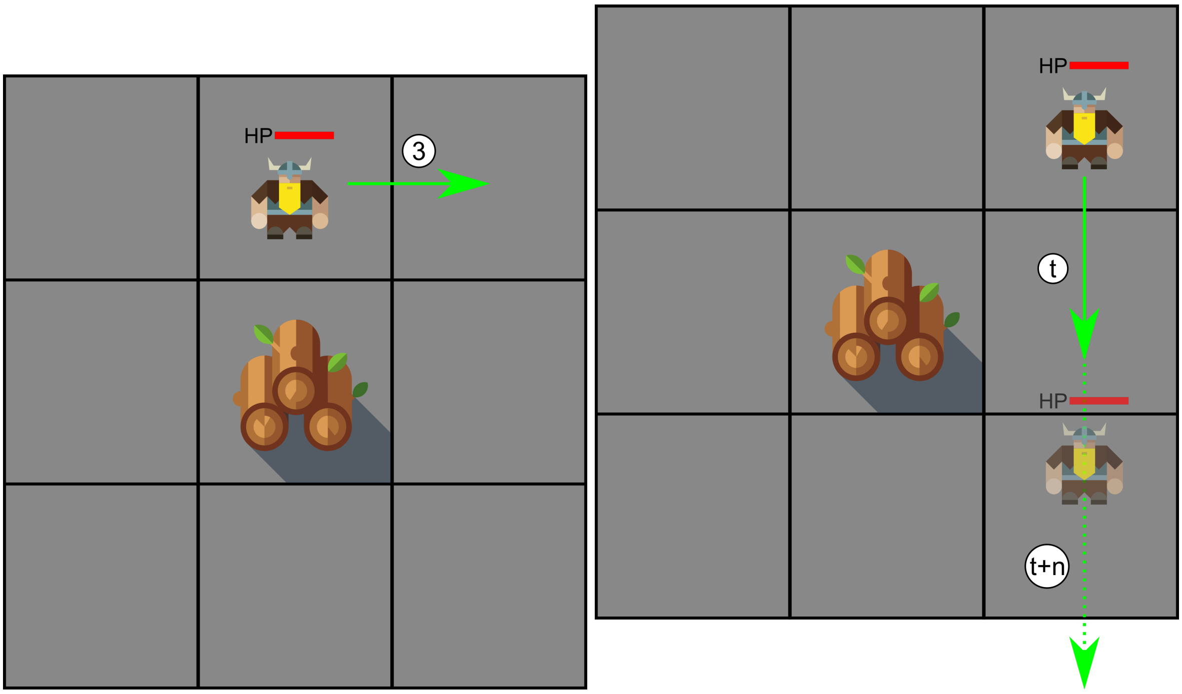


Figure 6.5 *Bar chart representing the average life span of a hero for each of the four controllers: DQN, A2C,* ***PrimEmo****, and a controller choosing a random action on each cycle. The error bar represent the* 95% *confidence interval.*



(a) Regardless of its environment, the hero first moves in one direction (left in this case).

(b) On the next game cycle, the hero moves back in the opposite direction.



(c) This loop repeats for tens of game cycles.

(d) Before the hero is again able to move in a more reasonable fashion.

Figure 6.6 A storyboard illustrating the oscillatory behavior displayed by the DQN controller.

Table 6.2 This table details the statistics describing the total reward each controller received during game.

| | mean | count | std | min | max | ci95_hi | ci95_lo |
|---|---|---|---|---|---|---|---|
| ctrl | | | | | | | |
| A2C | 26.220 | 400 | 39.710 | -61.349 | 151.951 | 30.112 | 22.329 |
| DQN | -16.036 | 400 | 19.489 | -96.437 | 10.801 | -14.126 | -17.946 |
| PrimEmo | -6.438 | 400 | 8.238 | -27.838 | 34.070 | -5.631 | -7.246 |
| Random | -7.222 | 400 | 5.214 | -15.039 | 18.265 | -6.711 | -7.732 |

Table 6.3 This table details the statistics describing the number of fights each controller has been involved in during a game.

| | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ctrl | | | | | |
| A2C | 0.009 | 130485 | 0.097 | 0.010 | 0.009 |
| DQN | 0.572 | 287800 | 1.401 | 0.577 | 0.567 |
| PrimEmo | 0.640 | 131574 | 1.076 | 0.646 | 0.634 |
| Random | 1.970 | 56003 | 2.932 | 1.994 | 1.946 |

better than random, but still very poorly. However, the videos showing a typical game for each of the controller type [5], tell a different story. From these it appears that the DQN controller either has not had the time to learn a complete strategy, or has already overwritten part of its initial strategy because it did not have enough neural units to store it in full. Either way, it seems that the only reason behind the DQN's longevity is that it displays oscillatory behaviors. At some point in the game, the hero controlled by the DQN algorithm moves back and forth between the same two tiles for hundreds of cycles before continuing (see Figure 6.6 for a representation of the DQN's oscillatory behavior). The rest of the time it manages to mostly stay out of harms way, but does not optimize its path toward the closest gold item or even try to avoid fighting enemies. As a result, even though the DQN has a longer life span than any other controllers, its overall performance shown in Figure 6.7, and in more details in Table 6.2, ends up being worse than that of the random controller. Consequently, the remainder of this section will focus solely on the results from the A2C and **PrimEmo** architectures.

---

[5] see: https://youtu.be/sG_weRLOhZQ for **PrimEmo**, https://youtu.be/RHE2DJM1MCA for the DQN, https://youtu.be/_8M6KnuJgSM for the A2C and finally https://youtu.be/rekPM-9K3sI for the random controller, as a comparison

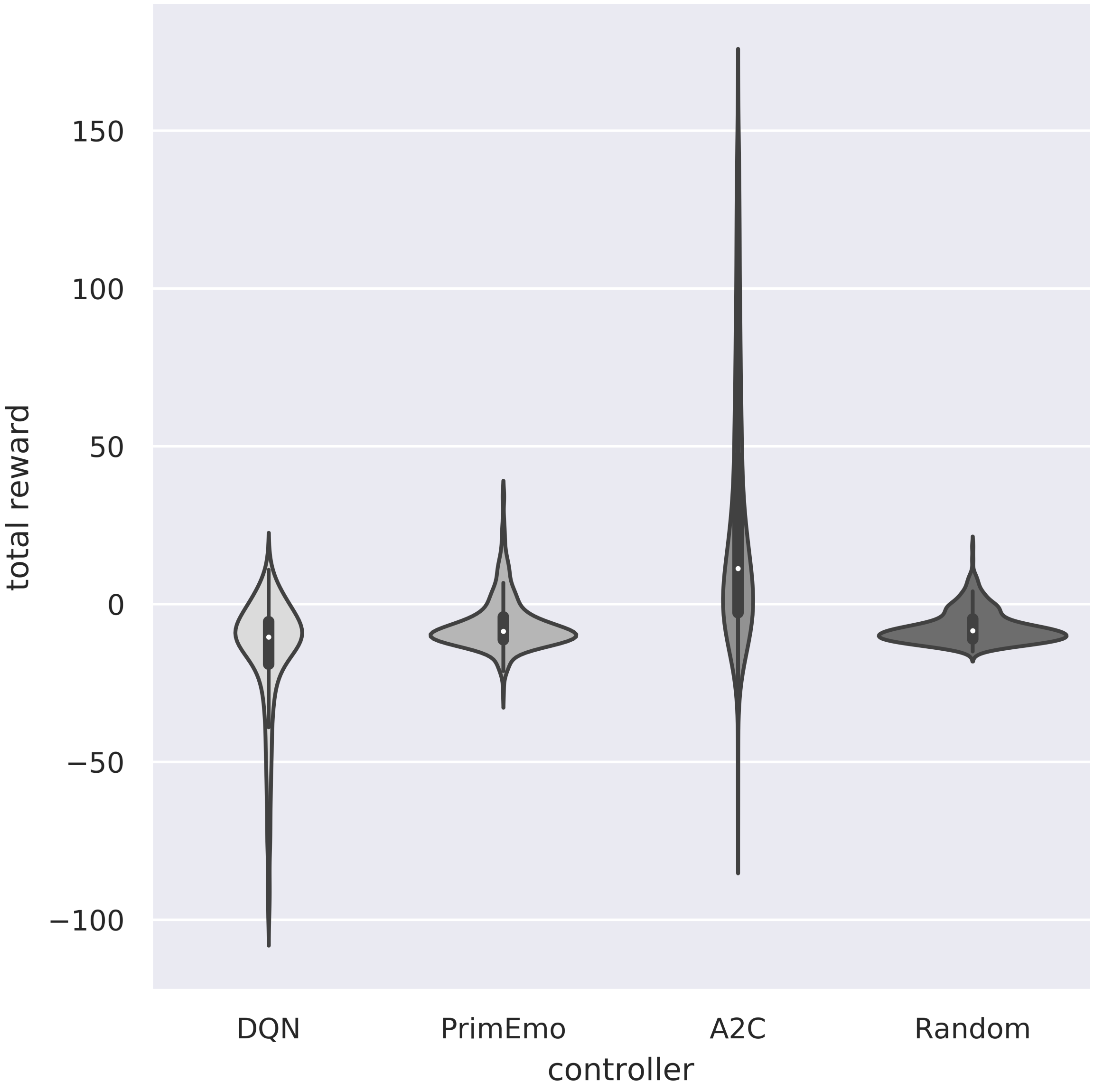


Figure 6.7 *Violin plot representing both the average total reward for each of the four controller, as well as the distribution of the total rewards for all the games played.*

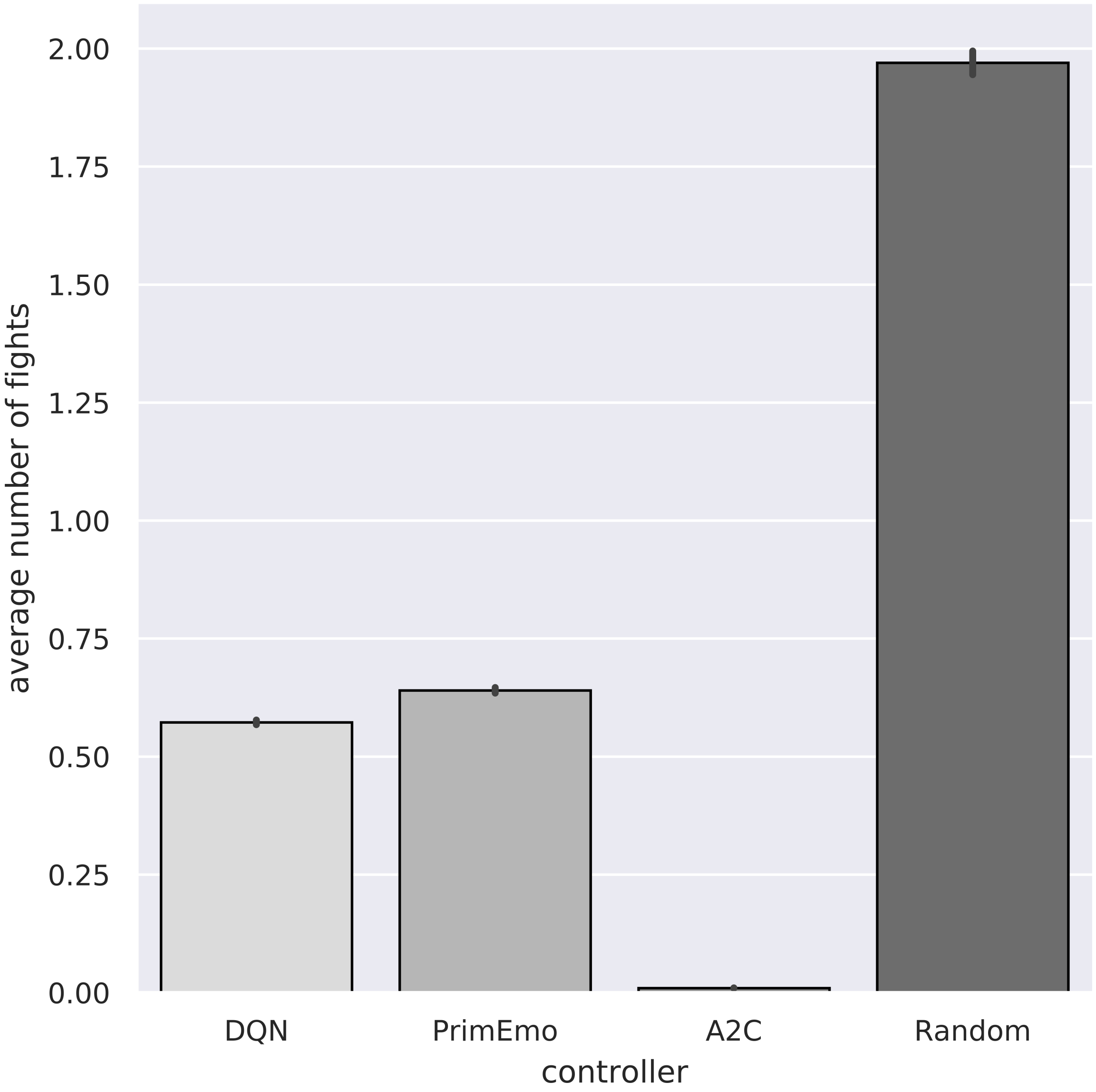


Figure 6.8 *Graph displaying the average number of time a hero fights during a game, for each of the four controllers. The error bar represent the* 95% *confidence interval.*

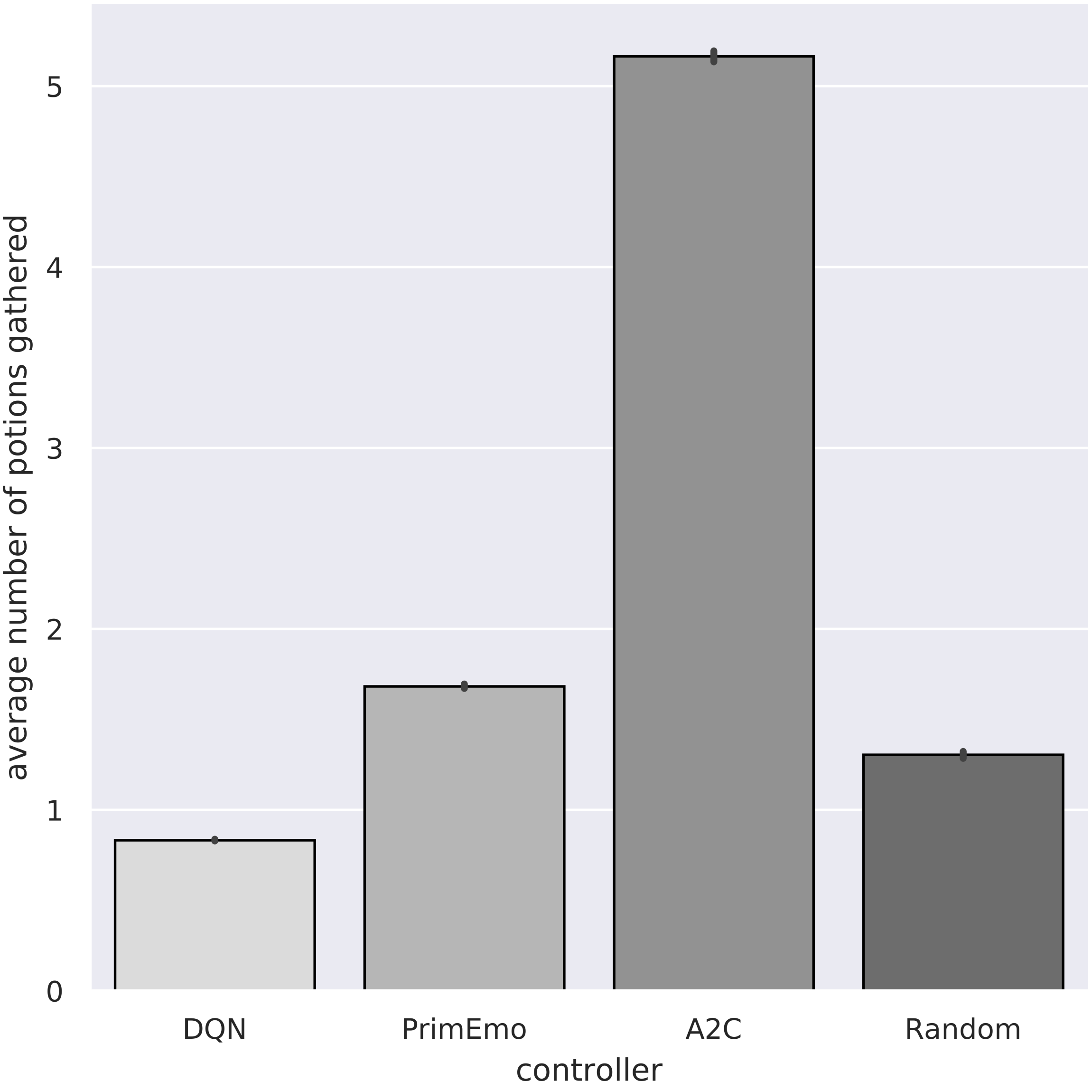


Figure 6.9 *Diagram displaying the average number of potions each controller picks up during a game. The error bar represent the* 95% *confidence interval.*

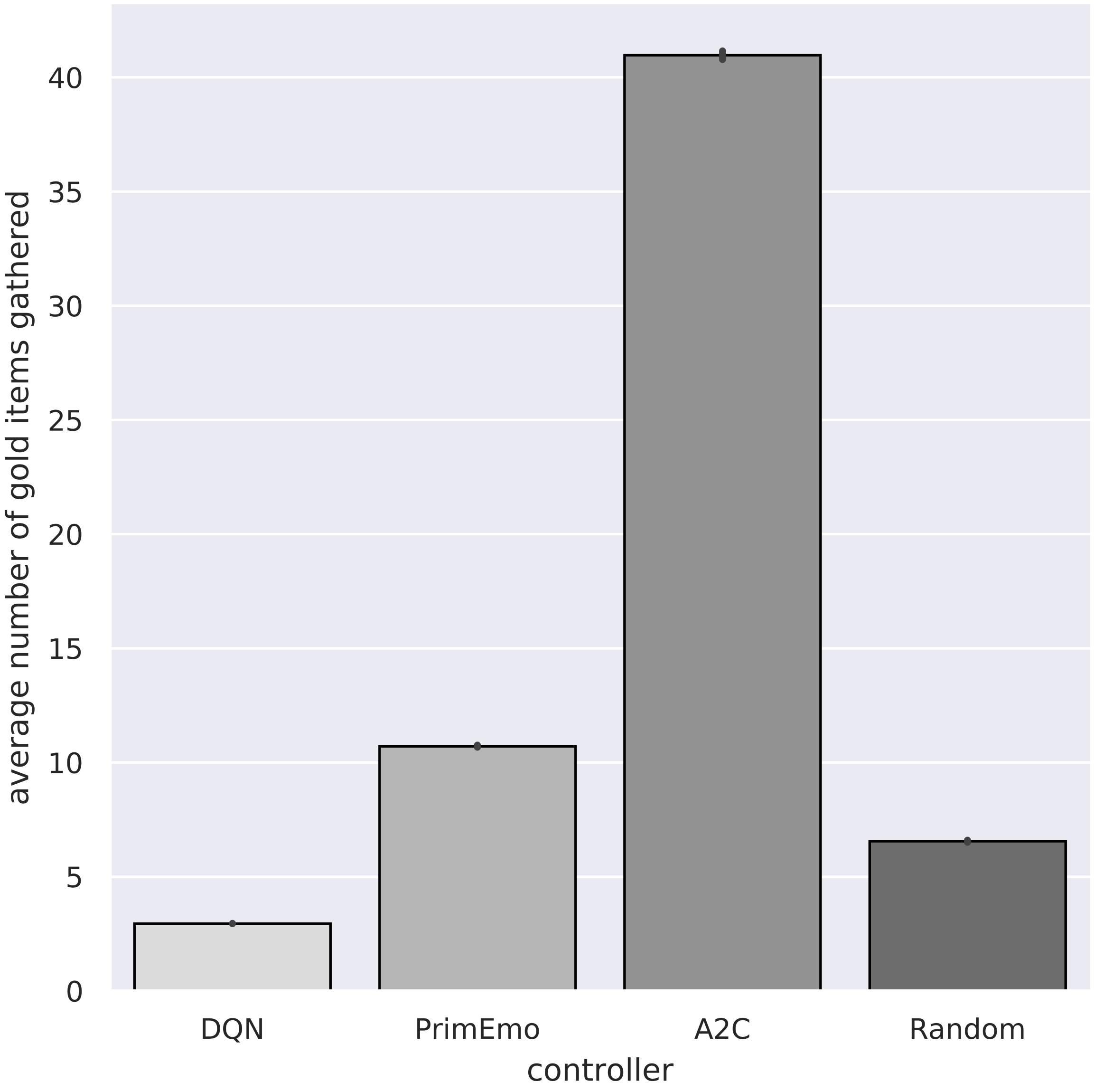


Figure 6.10 *Chart depicting the average number of gold chests each controller collects within a game. The error bar represent the* 95% *confidence interval. To win a game in Viking Doom a hero has to amass as much gold as possible, while staying alive.*

Table 6.4 This table details the statistics describing the number of potions each controller gathered during a game.

| | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ctrl | | | | | |
| A2C | 5.164 | 130485 | 5.552 | 5.194 | 5.134 |
| DQN | 0.833 | 287800 | 1.288 | 0.837 | 0.828 |
| PrimEmo | 1.683 | 131574 | 2.251 | 1.695 | 1.671 |
| Random | 1.304 | 56003 | 2.255 | 1.323 | 1.286 |

Table 6.5 This table details the statistics describing the number of gold items each controller gathered during a game.

| | mean | count | std | ci95_hi | ci95_lo |
|---|---|---|---|---|---|
| ctrl | | | | | |
| A2C | 40.970 | 130485 | 34.875 | 41.159 | 40.780 |
| DQN | 2.944 | 287800 | 2.686 | 2.954 | 2.935 |
| PrimEmo | 10.711 | 131574 | 8.658 | 10.758 | 10.664 |
| Random | 6.551 | 56003 | 6.146 | 6.602 | 6.500 |

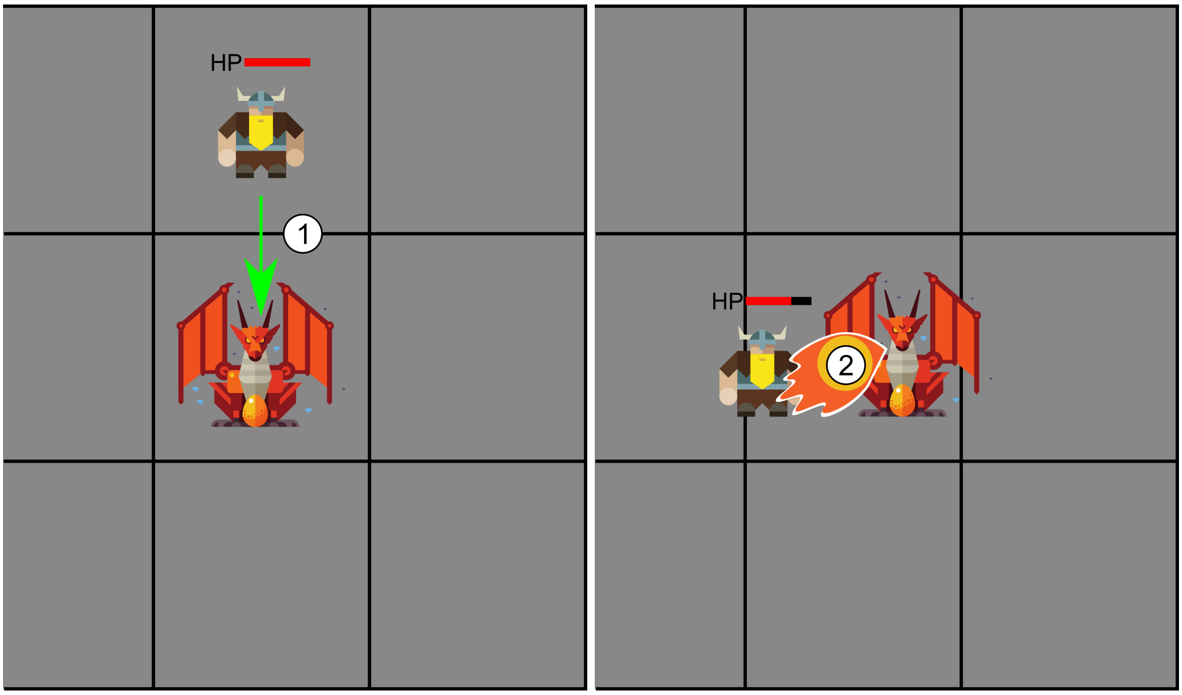
By analyzing the videos mentioned above, as well as Figures 6.8 — 6.10 and Tables 6.3 — 6.5, it is interesting to see that **PrimEmo** and the A2C controller have adopted opposite strategies. Indeed, on the one hand, A2C seems to be focused on gathering as much gold as possible and as quickly as possible, but at the cost of its hero's life. This is best illustrated by the fact that, although the A2C almost never fights (as shown in Figure 6.8 and Table 6.3), it is by far the controller that collects the most potion on average per game (see Figure 6.9 and Table 6.4). This might be surprising at first. However, a typical game cycle has the heroes act before the enemy units. This is so that heroes have the possibility to run away from a fight. What the A2C is doing in this case is quite the opposite. Since dragon units are moving at half the speed of a hero, whenever one of them is blocking the path to the nearest chest, rather than going around, the A2C chooses to step on the same tile as the enemy unit. As a result, the enemy performs an attack on the same turn, damaging the hero. On the next turn, the hero simply runs away in the direction of the closest chest without ever fighting (see Figure 6.11 for an illustration of the A2C controller's strategy). Once the hero's health reaches a certain level, the A2C simply collects one or more potion items before moving on to the next gold item (see Figure 6.13 for a depiction of this behavior). It is quite an effective strategy, since the A2C manages to place first not only in terms of gold gathered, but also in terms of average total reward per game, even if it is at the expense of the hero's well being.

On the other hand, **PrimEmo** seem to favor avoiding enemy units over collecting gold items. It is important to note that, given the difference in processing speed (clearly apparent in the videos) **PrimEmo** has had less time to refine its strategy and still shows signs of moving randomly at times. This is reinforced by Figure 6.7 and Table 6.2, which shows that **PrimEmo**'s overall performance is similar to that of the random controller. Pieces of a strategy can still be observed, since whenever **PrimEmo** finds itself near an enemy it tries to avoid it at all costs as depicted in Figure 6.12. If a battle is inevitable and the hero is damaged, **PrimEmo** will instruct it to move toward and collect the closest potion, to replenish its life as illustrated in Figure 6.13.

Even though, **PrimEmo** did not manage to learn as much as A2C, it is already quite an achievement for **PrimEmo** to be able to reach an average life span similar to heroes controlled by the A2C algorithm, which at the time of this implementation was considered state-of-the-art in reinforcement learning. Regarding Hypothesis 2-1, as disappointing as it may be, this means that it is impossible for the moment to give any definitive answer, since **PrimEmo** did not significantly outperform both DQN and A2C, but was not defeated either. However, given more time to optimize its strategy, it is certainly possible that **PrimEmo** could improve its overall performance, by gathering more gold and extending its hero's average life span.

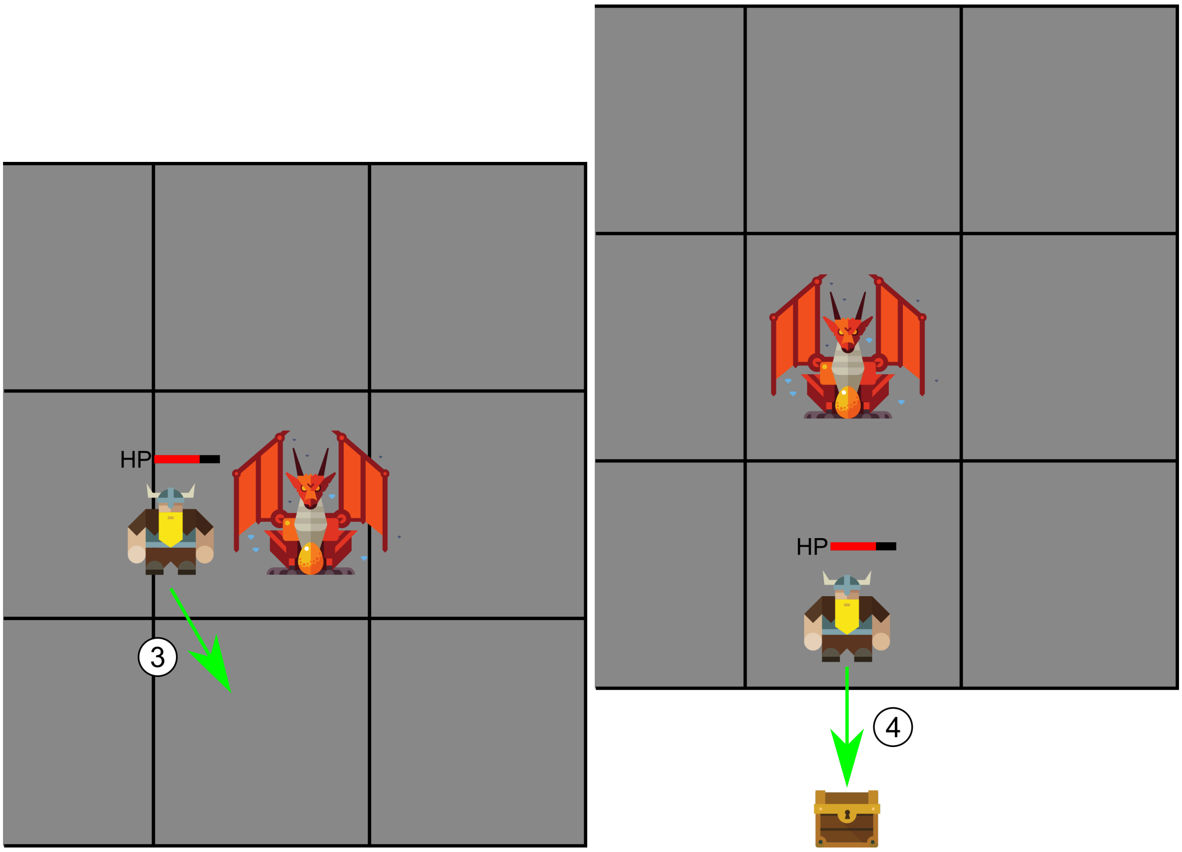
Concentrating on Hypothesis 2-2, it could be assumed that the conclusions drawn in Section 5.3 remain valid and, thus, only the second half of the hypothesis needs to be tested. However, this assumption is unsustainable because the context in which the amygdala operates has changed drastically. As shown in Figure 6.1, the amygdala receives input connections not only from the hypothalamus and thalamus, but also from the different cortical areas (from the somatosensory cortex to the central nucleus, and from both the pre-frontal and the motor cortices to the basal nucleus). As such, the information available to the amygdala is much more complex in **PrimEmo**, than it was in **ProtoEmo**. Therefore, it needs to be determined whether or not the amygdala is still responsible for detecting the salient features of a given situation and, hence, in charge of the '*arousal*' portion of emotions.

According to the top row of Figure 6.14 showing a polynomial regression (order=3) between the average activity of the central nucleus (CA) layer and the different input features, it appears that for at least the first four columns (concerned with the distances between the hero, the enemy



(a) The hero deliberately moves on the same tile as the dragon.

(b) On the same game cycle, the dragon attacks.



(c) On the next game cycle, the hero runs away.

(d) While the dragon is idle, the hero moves toward the gold chest.

Figure 6.11 A storyboard describing the main steps in the strategy implemented by the A2C controller.

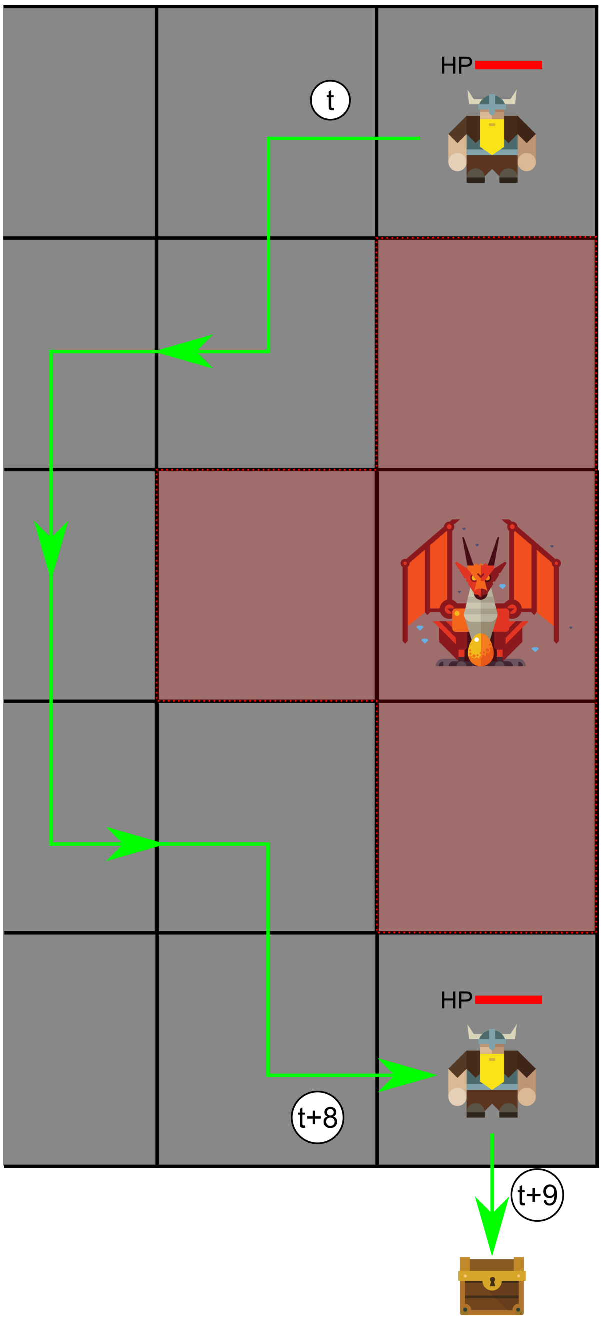
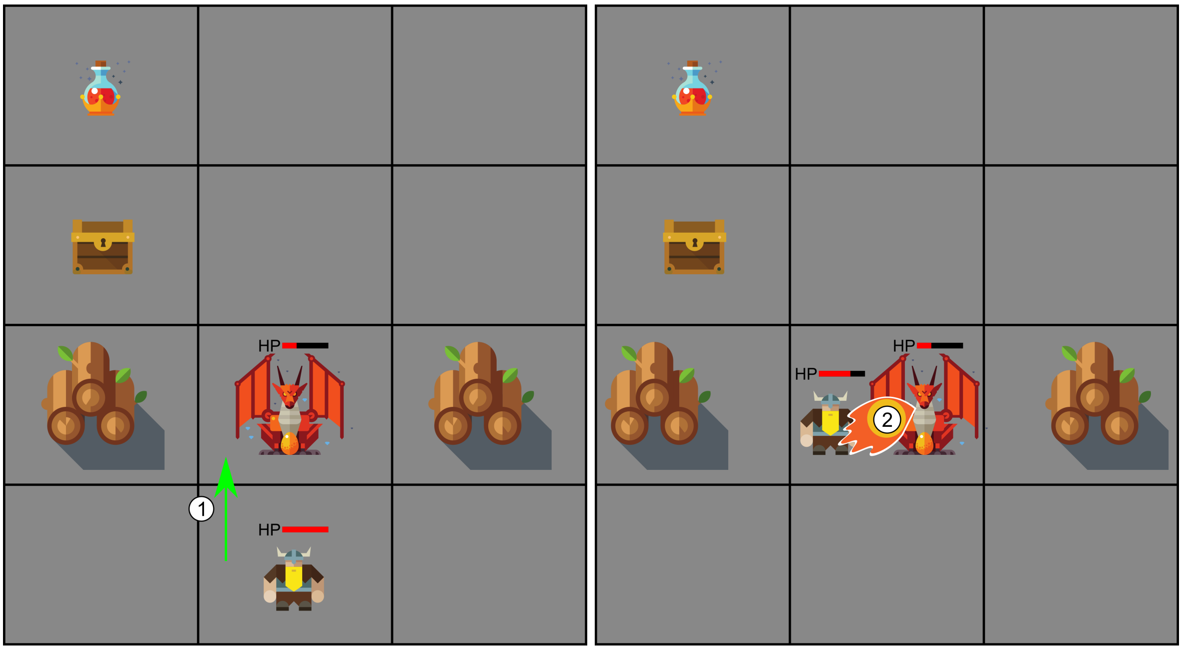
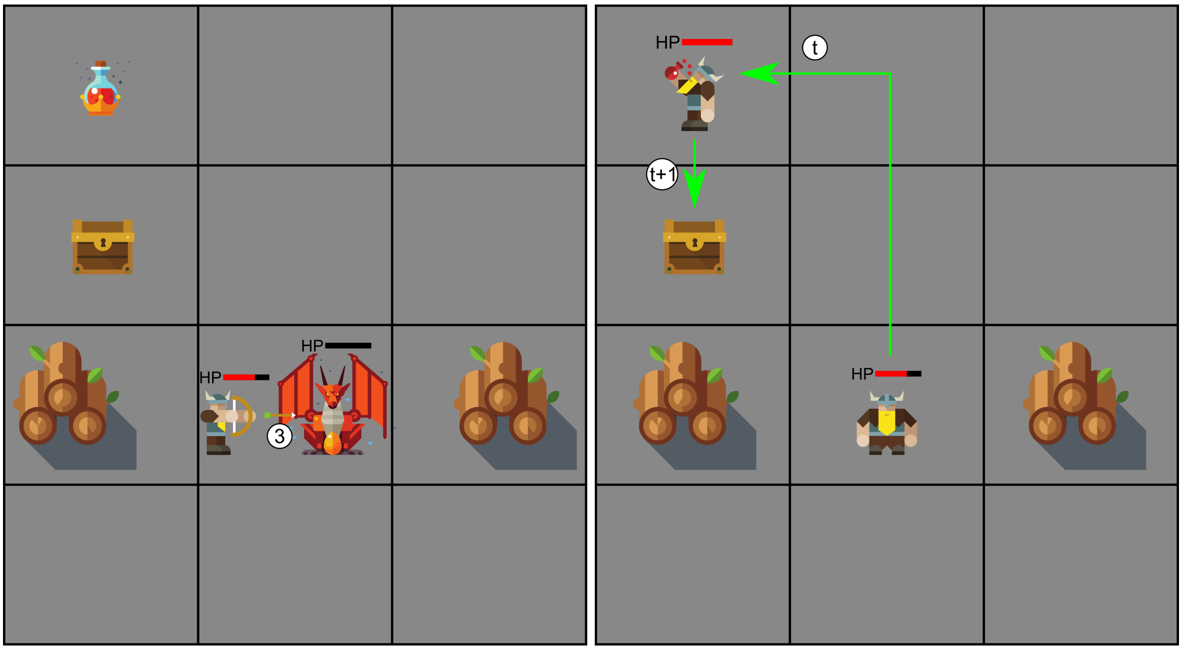


Figure 6.12 A figure depicting the strategy employed by the **PrimEmo** architecture, when it comes to dealing with an enemy.



(a) The hero deliberately moves on the same tile as the dragon.

(b) On the same game cycle, the dragon attacks.



(c) On the next game cycle, the hero defeats the dragon.

(d) The hero first collects the potion item to replenish its health, then moves toward the gold chest.

Figure 6.13 A storyboard representing a common strategy adopted by both the A2C and **PrimEmo** controllers. When the hero's health drops under a certain threshold, both controllers seem to prioritize collecting health items over gold ones.

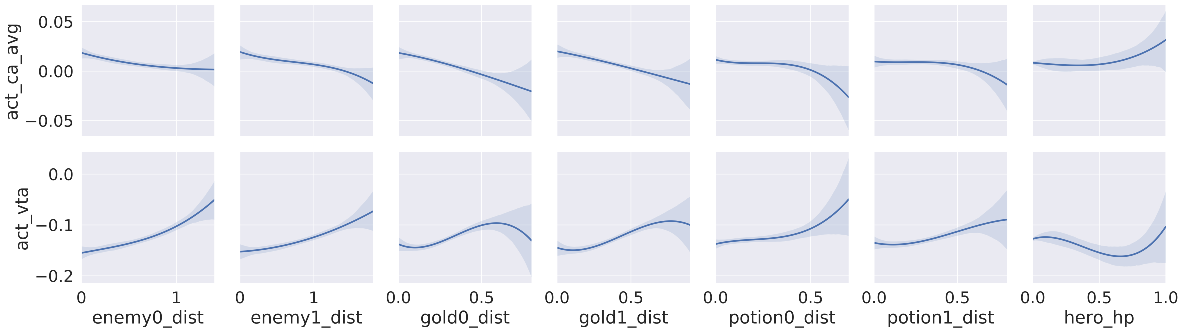


Figure 6.14 *Diagram showing the different polynomial regressions (order* $= 3$*) between the activities of the amygdala's central nucleus (CA), the ventral tegmental area (VTA) and the different input features: the distance to the two nearest enemy units, the distance to the two nearest gold items, the distance to the two nearest health potions, and finally the current hero's health level. The shadow around each line plot represent the* 95% *confidence interval.*

units and the gold items) the amygdala is indeed capable of detecting information relevant to both the hero's survival and its goal. Concerning the top row of the fifth and sixth columns, showing the evolution of the average CA activity against the distances to the two closest potion items, the negative correlation does not seem to be as strong. This is surprising, since potions being related to the hero's health should be quite valuable and, therefore, should trigger a burst of activity within the CA layer the closer the hero is to them. This could be explained by either **PrimEmo** needing a longer learning phase, or the data from the last column, which focus on the hero's health level. In this last facet, the CA layer is shown to have its highest activity, whenever the hero is at full life, while a decrease in health is accompanied by a decrease in activity. This is surprising not only because it means that **PrimEmo** does not value the life of its hero very highly, but also because it goes against the strategy extrapolated from the video and highlighted above. However, if the hero's health is indeed not a priority for **PrimEmo**, then it would make sense for the amygdala to not react to the potion items. The top row's last diagram, might also help shed some light on the results concerning the ventral tegmental area. With respect to Hypothesis 2-2 the amygdala was not capable of fulfilling its role. Moreover, it only aroused the rest of the architecture for some of the features relevant to the hero's survival, failing to detect the most important one: the level of health.

Given what is already known about the Ventral Tegmental Area (VTA) in animals (see Chapter 3), it is expected of the VTA layer to have a similar function within **PrimEmo**. More precisely, if the VTA layer is to predict the error in expected reward, then it should be positively correlated to the distance between the hero and the two nearest enemies, and to the level of health, meaning that the closer **PrimEmo** is to an enemy, the lower the reward. The same holds true for the evolution of the reward in regard to the hero's health.

In contrast, the VTA's activation value should be negatively correlated to the distance between the hero and both the two closest potions and the two closest gold items. Hence, the closer a hero gets to an item the more appealing the state. However, the results of the polynomial regression (order=3) in the bottom row of Figure 6.14 do not fully match those predictions. As highlighted above it appears that, during the learning phase, **PrimEmo** focused mainly on preserving the hero's life. As a result, the VTA's activity is clearly decreasing the closer a hero gets to an enemy unit. Remarkably, however, the VTA's activation values are also positively correlated to the distances between the hero and both the gold and potion items. It is especially surprising for the potions, since this would indicate that heroes should be running away from health items, while in the video **PrimEmo** was able to collect many of them without too much trouble. For the gold items, on the other hand, it further supports the fact that **PrimEmo** has not had the time to optimize its strategy. It also serves as an explanation for the random behaviors observed when a hero was near a chest. Since the CA layer only activates when the hero is at full health or far away from any potion items, as discussed above, it follows that the VTA would only activate and learn from those same situations. Hence, the polynomial regressions for the average VTA activation value, are only a consequence of activity of the CA layer. Consequently, because it was concluded earlier that the amygdala did not function as expected, it is only natural for the VTA layer to perform poorly as well. In regard to Hypothesis 2-2, this means that the VTA could not fulfill its role either.

Unfortunately, in its current state **PrimEmo** cannot trigger an emotional episode. Nevertheless, taking into account the fact that none of the controllers were able to complete their learning phase, it is impressive for **PrimEmo** to perform so well already, especially when compared to the A2C algorithm. It is, therefore, safe to speculate that, given more time to refine its strategy, **PrimEmo** would probably outperform the A2C controller. In so doing, both the amygdala and ventral tegmental area would also learn to perform their respective function more appropriately. Hence, validating Hypothesis 2-2 in the process.

# Summary

*To give it power over the decision-making process, and further increase its survival capabilities, the* ***ProtoEmo*** *architecture described in Chapter 5 has been integrated with O'Reilly et al.'s (2007)* 'Primary Value Learned Value' *mechanism and with the basal ganglia system, responsible for decision-making in the brain (see Chapter 3), resulting in the* ***PrimEmo*** *architecture. Given the increase in complexity and available information for the amygdala, I hypothesized that:*

**H2-1** *The use of a system capable of encoding the salience of survival relevant situational features will enhance the survival capabilities of virtual agents in complex and dynamic environments.*

**H2-2** *Furthermore, the mechanisms responsible for representing the salience of the survival relevant situational features will trigger an emotional episode. Emotions are differentiated from other mental states, by their level* 'arousal' *and* 'valence' *(see Sub-Section 2.7.3 for the exact definition of emotion adopted in this thesis).*

*To test the performances of the* ***PrimEmo*** *architecture in complex and dynamic environments, as well as explore the validity of the different hypotheses, a game called* 'Viking Doom' *has been created. Furthermore, two other controllers have been introduced as a comparison. The first one uses a* 'Deep Q-Learning' *(DQN) algorithm, while the second one implements an* 'Advantage Actor-Critic' *(A2C) model. In* 'Viking Doom' *each controller is responsible for a hero, which can move in four directions (up, down, left and right) or stay idle. The goal for all heroes is to gather as much gold as possible before the end of the game, while staying alive. To do that, the different controllers have to build a strategy optimized for collecting gold items and avoiding as much as possible any fight with enemy units. Collecting items or fighting with enemies happens automatically, whenever a hero finds itself on the same tile as an item or an enemy unit, respectively. From the results gathered during this experiment, it was clear that none of the controllers were able to learn a complete strategy. This is best exemplified by the DQN controller, which displayed oscillatory behaviors during its testing phase. Focusing on* ***PrimEmo*** *and the A2C algorithm, although the latter had a much better overall performance, both controllers had similar average life spans. This means that because* ***PrimEmo*** *did not outperform A2C in terms of survival, hypothesis H*2-1 *was deemed invalid for the moment. However, given that* ***PrimEmo*** *had less time to refine its strategy and that it still performed comparably to the A2C model, it is suspected that with some improvement hypothesis H*2-1 *could be verified. Finally, even though the average overall activity for the amygdala's central nucleus had the right correlations with the input features, the layer corresponding to the ventral tegmental area was not able to learn the correct values for the different situations encountered by the hero. As a result, hypothesis H*2-2 *was invalidated as well.*

# CHAPTER 7

## Conclusions

In conclusion to this thesis, this last chapter takes a look back at this project. First synthesizing its content, to then be able to clearly answer the research hypothesis, formulated in the introductory chapter, that guided this endeavor. It then explores the different contributions brought forth by the experimental results gathered and architectures designed during this journey. Finally, possible future extensions to the work presented in this thesis are highlighted, before the concluding statement.

### 7.1 Scope of this thesis

The overall goal of this project was to understand the role of emotions in the decision-making process, as well as investigate the brain regions involved. Since it is believed that emotions are widespread among the animal kingdom, and because nature usually does not like to waste energy on non beneficial systems, it means that emotions have some evolutionary advantage. Otherwise, they would be a much rarer trait or even extinct already.

However, before even beginning to think about the influence emotions have over the rest of the brain, it is important to survey the field of affective science to understand what sets an emotion apart from any other neural activity. Chapter 2, therefore, covered the history of emotions. How they were conceptualized, from ancient Greece, to Descartes and finally to today. Three school of thoughts were explored in more details, because of their current prominence. Section 2.4 focused on the appraisal theory, which view emotions as state of mind labeled by a cognitive system. Using the tools afforded by mathematics, especially statistics, dimensional theories usually consider emotions to be a point in a dimensional space. Section 2.5 detailed some of the dimensions that have been used to characterize emotions. Amongst the more popular ones, Mehrabian's (1980; 1996) Pleasure-Arousal-Dominance scale is the one that stood the test of time. The last sub-field surveyed, that of anatomical theories, sits at the intersection between affective science and neuroscience. Consequently, researchers in this field usually focus their investigation on the brain regions that might originate emotions. Section 2.6 gives a chronological overview of the many discoveries made since Darwin. Along the way some theoretical models of the origin of emotions within the brain were also studied. Models such as Papez's (1937) circuit or MacLean's (1970) triune brain (both described in Section 2.6) still influence our current interpretation of the brain and emotions within it. One model, though, differentiate itself from the rest. Extensively reviewed in Sub-Section 2.6.2, the '*Survival Circuit*' theory (advocated by LeDoux (2012)) reflects on the many criticisms made against the '*basic emotion*' theory (detailed in Sub-Section 2.6.1). Borrowing the idea that emotions evolved to improve survival, from Darwin (1872), it suggests to focus on the link between emotion and survival, rather than looking for a "*happiness circuit*" within the animal's brain. A survival circuit is, therefore, described as an innate neural architecture capable of detecting opportunities relevant to an individual's survival and of triggering innate reactions. LeDoux (2012) purposefully avoids defining the concept of emotion, instead relegating it to a simple label applied by our conscious mind whenever the triggering of a survival circuit is detected. This thesis takes inspiration from the defense circuit established in LeDoux (2012).

This overview of the landscape of affective science was followed by an exploration of the most popular or recent computational models of emotions hailing from the three main sub-fields. The reason for this second survey is that implemented models are grounded in reality, while theories often suggest abstract answers. Computational models are a realization of the chosen theory. Rather than painting a picture with broad strokes, the author has to be more precise and define every little details before being able to implement the model. It restricts the theory to a single interpretation, but at the same time it might also provide concrete answers to some of the questions left behind. Furthermore, this also gives an idea of the direction the field is headed toward, as well as the questions that remain unanswered. Finally, because it represents a single interpretation and due to the many design decisions made during the building process, usually only the most relevant or core features of a given theory are actually embedded in the final model. As such, it is easier to understand what are the recurring features used to characterize emotions across the different sub-fields. From the surveyed models, we were able to derive that the most consistently used qualities to discriminate emotions from other mental states seem to be ‘*valence*’ and ‘*arousal*’. Borrowing the vocabulary of dimensional theories, ‘*arousal*’ describes the level of physiological activity associate with an emotion, while ‘*valence*’ denotes either a pleasurable or non-pleasurable feeling accompanying an emotion.

Next, the **ProtoEmo** architecture was designed to both validate the chosen definition of emotions and show that the circuit implemented by the thalamus, hypothalamus and amygdala is capable of encoding the saliences of survival relevant situational sensory stimuli. The importance of this seemingly small architecture, cannot be overstated since it stands as the beginning of an emotional episode. Indeed, assuming a role akin to an alarm system, **ProtoEmo** is capable of warning other brain areas of the presence of relevant stimuli. Where the relevance of a given stimulus is relative to the individual’s believes, desires and intentions. This in turn allows the brain to better manage its limited cognitive resources and only focus on the important matters. In terms of emotions, **ProtoEmo** only implements an archaic version of the concept, since it can only influence the decision-making process and contributes to arousal of the brain. However, it has the advantage of knowing about both internal and external states, along with having a widespread influence.

Guided by the concept behind the survival circuits theory, suggested by LeDoux (2012), the next logical step was to explore the brain mechanisms dedicated to action selection. Indeed, according to the survival circuits theory, emotions evolved among the animal kingdom as a system to increase an individual’s chances of survival. As explained in more details in Chapter 2, the main way in which this benefit to survival manifests itself is via the direct control of behaviors. To put it more simply, it steers the animal toward rewarding stimuli and away from punishing situations. As detailed at length in Chapter 3, the underlying mechanisms theorized to signal reward and punishment within the brain are the dopamine and serotonin systems, respectively. Of the two, the dopamine system has been the center of attention and its activity has been mathematically modeled. Thanks to the work of Sutton and Barto (1998) it is in fact the basis for a well-known paradigm called reinforcement learning. Within this paradigm, an agent will simply learn to solve a task through a trial-and-error process, in which the environment provides rewards and punishments to guide the agent toward the goal. The dopamine system, though, is only the Critic part of the decision-making process. According to both Redgrave et al. (1999) and Frank et al. (2001) the part of the brain that actually learns the policy and selects the next action to perform are the basal ganglia. The basal ganglia implement a gating mechanism, which under the influence of the dopamine system's activity either allow or prohibit a single action to be executed. In this case the action can either be motor or cognitive in nature. This means that the basal ganglia, not only control outward behaviors, they also manage the access to cognitive resources by other control systems within the brain.

To explore the role of emotion in the decision-making process described in Chapter 3, along with the benefits afforded by the alarm system implemented by **ProtoEmo**, a second architecture was designed. This new architecture has been referred to as **PrimEmo**, since it is hypothesized to support a primitive form of emotions, characterized by their level of arousal and valence (as mentioned above). To ascertain the capabilities in terms of survival and decision-making of the **PrimEmo** architecture, an online roguelike game was implemented. In it an agent controlled by **PrimEmo** had to compete against other players, namely a Deep Q-learning agent and an Advantage Actor-Critic agent. The goal of the game, apart from surviving in a rather harsh environment (populated by enemies) was to collect as much gold as possible.

From the experimental results gather through this game, it was concluded that agents controlled by the **PrimEmo** architecture had similar performances in terms of average life span than those using the A2C controller. In contrast with the A2C controller, the agent under **PrimEmo**'s influence was more focused on preserving its life, than collecting gold. Unfortunately, since **PrimEmo** did not have enough time during the learning phase to refine its strategy, neither the amygdala layer nor the ventral tegmental area layer were able to behave as expected. Therefore, it is impossible to draw any conclusions regarding **PrimEmo** supporting a primitive form of emotions characterized by their valence and arousal.

## 7.2 Answers to the research questions

At the beginning of this thesis is was established that to make its scope more manageable, the overall research question should be narrowed to single cognitive system. Therefore, rather than exploring '*the role of emotions in autonomous social agents*' in general, the goal of this thesis is to understand the function emotions play in the decision-making process. In this context it was hypothesized that: "*Emotions inform the brain as to the nature of a given situation, interrupt any ongoing action, and guide the decision-making process to increase the chances of survival of virtual agents*".

After extracting the different features that make emotions unique within the brain, as well as exploring their role in improving an animal's survival potential and their neural underpinnings, it became apparent that the primary hypothesis needed to be broken down further if we ever hoped to answer it. Consequently, Chapter 5 sought to verify that:

**H**1-1 The circuit identified as important to animal's survival, can be used by virtual agents to influence their actions and increase their survival capabilities.

**H**1-2 Within the previously established circuit, the amygdala acts as the triggering mechanism, by encoding the salience of features relevant to the survival of the individual in a given situation.

**H**1-3 This same circuit, however, is not enough to elicit emotions as characterized at the end of Chapter 2.

To verify the veracity of these hypotheses, the **ProtoEmo** architecture was built. Its performances were tested on resource foraging task, against another type of emotional agent described by Scheutz (2004). From the results gathered during this experiment (detailed in Section 5.3), we extrapolated that while Scheutz's agents were implementing a strategy, were only the strongest survives, ProtoEmo had a fairer approach, in which agents would avoid fighting at all cost and would explore the environment if the nearest energy source was already being targeted by someone else. Consequently, not only did **ProtoEmo** have a higher number of survivors at the end of each simulation, but it also had a greater maximum population, on average. As a result, it was concluded that in this case the Hypothesis 1-1 was verified. Further investigating the correlation between the value of the tendencies and the different input features, showed that the central nucleus of the amygdala behaved within expectations. In other words, the amygdala was able to detect the different opportunities relevant to the agent's survival, while avoiding any potential danger or battle with other agents. Hence, confirming the validity of Hypothesis 1-2. However, even if the correlation between the average activity of the amygdala's output layer and the input features, revealed that the closer an agent was to either danger or an energy source, the higher the activation rate, the same signal did not carry any information concerning the valence of the situation. Therefore, Hypothesis 1-3 was considered true as well.

Building on the knowledge accumulated so far, Chapter 6 introduced **PrimEmo**, which extended the **ProtoEmo** architecture with neural mechanisms capable of making decisions (see Chapter 3 for more details on action selection). The goal, in building **PrimEmo** was to substantiate the following hypotheses:

**H**2-1 The addition of a system capable of encoding the salience of situational features relevant to the survival of the individual will enhance the survival capabilities of virtual agents in complex and dynamic environments.

**H**2-2 Furthermore, the mechanisms responsible for representing the salience of survival relevant situational features will trigger an emotional episode, where emotions are differentiated from other mental state by their level of ‘*arousal*’ and ‘*valence*’.

In lieu of a complex and dynamic environment, an on-line roguelike game was implemented, so that **PrimEmo**’s performances could be compared to that of a Deep Q-Learning (DQN) and an Advantage Actor-Critic (A2C) algorithms. After collecting the results from this experiment and watching the controllers at play (footage of typical games are available on-line for each controller [1].), it is apparent that none of them were able to complete their training phase. Nevertheless, the data was analyzed to draw tentative conclusions regarding the hypotheses. Given that the DQN controller displayed oscillatory behaviors and performed worse than the random controller, it was left out of the comparison. In terms of average life span, both **PrimEmo** and the A2C algorithm performed similarly. Concerning their respective strategies, though, they could not be further apart from each other. While the A2C is clearly focused on gathering as much gold as possible, at the cost of its agent’s life, **PrimEmo** seems to favor avoiding enemy units as much as possible. However, because **PrimEmo** did not significantly outperform the A2C controller, it was concluded that a definitive answer to Hypothesis 2-1 could not be given. Still, the fact that **PrimEmo** was able to perform as well as the A2C controller, and that its strategy seems focused on preserving its agent’s life above all, seems to indicate that, given more time to refine its strategy, **PrimEmo** could indeed enhance the survival capabilities of virtual agents. For the second hypothesis (H2-2), the conclusion was even worse. The correlation between the average amygdala activity and the different input features made it clear, that, although, the amygdala was able to detect danger coming from nearby enemy units, it was unable to recognize, when the agent’s level of health was alarmingly low. Similarly, the ventral tegmental area, suspected to be in charge of the ‘*valence*’ of emotions, rewarded the agent for being far from any enemy units, but also for having a low level of energy. Consequently, in its current state the **PrimEmo** architecture could not verify the second hypothesis.

In view of the problems encountered in Section 6.4, it is unfortunately not possible to give any conclusive answer to the main hypothesis. The results from the first simulation

---

[1] `https://youtu.be/sG_weRLOhZQ` for **PrimEmo**, `https://youtu.be/RHE2DJM1MCA` for the DQN, `https://youtu.be/_8M6KnuJgSM` for the A2C and finally `https://youtu.be/rekPM-9K3sI` for the random controller

indicate that the survival circuit identified by LeDoux (2012) are indeed capable of increasing the survival potential of virtual agents. Although, the integration between this circuit and the decision-making process needs further investigation it still shows a lot of promise. In conclusion, therefore, even if the main hypothesis has not been verified yet, allowing the **PrimEmo** architecture to further refine its strategy could see it succeed in confirming both of its associated hypotheses. In turn, this thesis would have been able to show that: "*Emotions inform the brain as to the nature of a given situation, interrupt any ongoing action, and guide the decision-making process to increase the chances of survival of virtual agents*".

## 7.3 Future Work

Even though the work presented in this thesis is restricted to the study of the role of emotions within a single cognitive system, the ideas and methodology followed should act as a trigger for a host of research projects and discoveries in both cognitive and affective science. As such, the extensions highlighted in this section represent a possible trajectory between the results presented in this thesis and fulfilling the goal of both AGI and cognitive science.

Before any further experimentation occur two points should be addressed. To begin with, as pointed out in Sub-Section B.1.1 the neural units used for the actual implementation of the **PrimEmo** controller all applied a '*Softsign*' activation function to the weighted sum of their input activities. This is an error on my part, since the softsign function was meant to be used for the neurons in the ventral tegmental area (VTA) only. Indeed, as argued in Sub-Section A.1.1 a ReLU activation function is enough and in this case better suited to model the behavior of cortical neurons. Furthermore, the reason for using a softsign activation function for neurons in the VTA is that modulation within the **PrimEmo** controller has been implemented as a product between the VTA's output activity and the activation values of the neurons part of the layer receiving projections from the dopaminergic system (see Sub-Section B.1.1 for more details). Since a single layer was used to represent the activity of both the dopaminergic and serotonergic systems (as explained in Section 6.1), and because the values of softsign are in the range $[-1,1]$, it means that a VTA output of $-1$ is interpreted as a punishment, a value of 1 would be a reward, and 0 is considered neutral. The issue with the current implementation of modulation is that, if the VTA layer sends a neutral signal to any of its targeted areas, then the neurons of the receiving layer would be completely silenced. To solve this problem two solutions are available. The first is to simply add a constant to the output activity of the VTA layer, so that 1 would be considered the neutral signal, 0 a punishment, and anything above

1 would be interpreted as a reward. Another possibility is to implement modulation as an additive operation as suggested by Litt et al. (2006) (and supported by Frank (2005) and Gurney et al. (2001b)). Adopting this solution, the output activity of a neuron targeted by both the dopaminergic and serotonergic systems is expressed as:

$$a(t) = a_0(t) + \beta \times DA(t) + \gamma \times 5 - HT(t) \tag{7.1}$$

Where $a_0(t)$ represent the neuron's activity at time $t$ before modulation, $DA(T)$ is the output value of the dopaminergic system, and $5 - HT(t)$ is the output activity of the serotonergic system at time $t$.

The next step, then, before looking into the future should be to adjust the different controllers' hyper-parameters, including those related to **PrimEmo**, until all controllers are able to play Viking Doom in its complete form (with all items, enemies, obstacles, and heroes on the same map) to a reasonable level. Only then will it be possible to gather data relevant to the hypotheses described in Chapter 6, and analyze those results to draw definitive conclusions for both **PrimEmo** and the global hypothesis presented in Chapter 1.

Once such a baseline has been established, it would be interesting to explore the consequences of modifying some of the connections in **ProtoEmo**. As introduced in Section 6.1 and further detailed in Section B.1.2, in **PrimEmo** the hypothalamic layer (labeled '*Hypot*' in Figure B.1a and Sub-Figure B.1a) is bidirectionally connected with the basal nucleus of the amygdala (BA). While the projection in the direction hypothalamus to BA makes sense, since it provides the BA with a description of the internal state of the individual, the connection in the other direction (BA to hypothalamus) might not be required. Indeed, the link from BA to hypothalamus has been described as relatively weak (Whalen & Phelps, 2009). Furthermore, if the loop formed by the lateral hypothalamic area, the hypothalamus, the BA, and the CA is to be interpreted as an attention mechanism (as explained in Chapter 2 and Sub-Section A.1.4), in this context then the projection from the BA to the hypothalamus serves no purpose, save maybe redundancy or descending feedback (but see Sub-Section B.1.2 for another interpretation). Now that the BA receives information about the internal state of the individual directly from the hypothalamus, and given that the function of the BA has been described as contextualizing the states of both the individual and the environment using the content of working memory (see Sub-Section B.1.3), it would be interesting to investigate the usefulness of the projection from the hypothalamus directly to the central nucleus of the amygdala (CA). Indeed, in Chapter 5 the structure of the amygdala has been described as being similar to that of a funnel, where information entering the lateral nucleus of the amygdala (LA) is incrementally refined until it reaches the CA. At which point, the CA is only responsible for detecting stimuli relevant to the survival of the individual. Therefore, having a connection from the hypothalamus to the BA, as well as to the CA seem redundant, and could interfere with the function attributed to the CA. Removing the projection between the hypothalamus to the CA would mean that the BA integrates sensory information describing the state of both the environment and the individual with the content of working memory and funnels the result to the CA, which is then only responsible for alarming the rest of the brain whenever a stimulus of major concern is detected.

The lateral nucleus of the amygdala (LA) has been presented in Chapter 5 and Appendix A as the interface to the amygdala. In the implementation of the **PrimEmo** controller detailed in Appendix B, though, the LA only receives sensory related data from the thalamus. The somato-sensory cortex (SSC) instead connects directly to the central nucleus. Once again by-passing entirely the funneling structure described in Chapter 5. Therefore, for the design to be more biologically appropriate, and for it to more faithfully reproduce the '*dual pathways*' to the amygdala as suggested by LeDoux (1986, 1992) (see Chapter 4 for a complete description of this model), the SSC should send efferents targeting the LA rather than the CA.

Having experimented with and refined the **PrimEmo** architecture, it is finally time to look into the future and explore the possibilities this thesis created. A starting point on the path toward AGI, is the integration of **PrimEmo** into a cognitive architecture. Taking ACT-R (J. R. Anderson et al., 2004) as an example, **PrimEmo** would replace the decision-making module. Communicating via the different buffers to request and retrieve information necessary to its role. A well-known architecture, such as ACT-R or SOAR (Laird et al., 1987), would provide the perfect testbed for integrating a connectionist system within a symbolic one. The fact that those architectures are well-known, means that they have been and still are used for both scientific and industrial application. Consequently, there should be an abundance of empirical results from different tasks, depicting the capabilities of those cognitive architectures. A successful integration of **PrimEmo** within such systems, would be characterized by its ability to perform in a similar fashion or even better than the original architecture, on the same sets of tasks.

In a later phase, it would be beneficial to both cognitive and affective science, to replace or rebuild each module from ACT-R or any other cognitive architecture, with biologically plausible mechanisms. Taking care to always include both cognition and emotions as the focal point of the study. This would not only increase our current understanding of the brain's most basic systems, but also establish a strong and reliable basis upon which the next phase could expand. Furthermore, in a similar fashion than in the first step, there is still the advantage of having the original architecture to assess the capabilities of its biologically plausible counterpart. The reasoning behind this suggestion, is that a lot of thoughts have already been given to the cognitive systems that are necessary for a minimal functioning brain replica. Thus, investigating the underlying neural circuits, as well as the role emotions play in all those systems (similarly to what has been done in this thesis), seem like the most sensible next step towards AGI.

Once a reliable core architecture has been designed, it is only a matter of repeating the same process to each and every cognitive systems. Hence, incrementally extending the capabilities of the original construct, until it reaches a point where it is endowed with human-level intelligence. Unfortunately, the further the architecture develops, the more complex it will be to go on. This is because, as cognition and emotions reach higher levels, both systems become increasingly more intricate and so do their interactions (Alexandrov & Sams, 2005; Ortony et al., 2005). It is small wonder some think that humanity will never fully explain the concept of consciousness. If that is true, then the same could be said of full-fledged emotions. Currently, it feels like affective science is still lagging behind cognitive science. The hope is that by grounding the different theories of emotions in reality (through the implementation of modules for cognitive architecture) affective science should quickly catch up. From this point on both fields would then be co-evolving. Discoveries made in one, would inform the other and vice versa, creating a '*never*' ending positive feedback loop.

## 7.4 Conclusion

In the end, what this thesis really contributes to the edifice of cognitive and affective science, is not knowledge. Where knowledge is understood as a theory or model of emotions. Rather, it provides evidences that emotions can be investigated by relying on their roles, instead of their many aspects. Furthermore, it also offers a point of view where affect and cognition are two co-evolving and inter-dependent systems. Where the line between those two concepts is blurry. Consequently, reaching AGI will require many more projects of a similar nature. Each investigating the interactions between a given cognitive system and emotions. Each extending the architecture built by those before them, as well as our understanding of the brain and intelligence in general.

# References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.*

Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., … Sowa, J. (2012, mar). Mapping the Landscape of Human-Level Artificial General Intelligence. *AI Magazine*, *33*(1), 25.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005, jan). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72.

Aggleton, J. P. (1992). *The Amygdala : neurobiological aspects of emotion, memory and mental dysfunction*. New York: Wiley-Liss.

Alexandrov, Y. I., & Sams, M. E. (2005, oct). Emotion and consciousness: ends of a continuum. *Cognitive Brain Research*, *25*(2), 387–405.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., … Sobel, N. (2003, feb). Dissociated neural representations of intensity and valence in human olfaction. *Nature neuroscience*, *6*(2), 196–202.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Arbib, M. A. (2003). *The handbook of brain theory and neural networks* (2nd ed. ed.; M. A. Arbib & I. NetLibrary, Eds.). Cambridge, Mass.: Cambridge, Mass. : MIT Press, c2003.

Aristotle. (1926). *The Nicomachean Ethics* (H. Rackham, Trans.). Cambridge, MA: Harvard University Press. (Original work published 350 B.C)

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., & LeDoux, J. E. (1995, apr). An anatomically constrained neural network model of fear conditioning. *Behavioral Neuroscience*, *109*(2), 246–57.

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., & LeDoux, J. E. (1997, apr). Computational modeling of emotion: explorations through the anatomy and physiology of fear conditioning. *Trends in cognitive sciences*, *1*(1), 28–34.

Armony, J. L., Servan-Schreiber, D., Romanski, L. M., Cohen, J. D., & LeDoux, J. E. (1997, mar). Stimulus generalization of fear responses: effects of auditory cortex lesions in a computational model and in rats. *Cerebral cortex (New York, N.Y. : 1991)*, *7*(2), 157–65.

Arnold, M. B. (1959). From perception to emotion and action. *Acta Psychologica*, *15*, 406–417.

Arnold, M. B. (1961). *Emotion and personality*. London: London : Cassell, 1961.

Atallah, H. E., Frank, M. J., & O'Reilly, R. C. (2004, nov). Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. *Neurobiology of learning and memory*, *82*(3), 253–67.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Progress in Brain Research*, *150*, 45–53.

Bard, P. (1929, aug). The central representation of the sympathetic system: As indicated by certain physiologic observations. *Archives of Neurology And Psychiatry*, *22*(2), 230–246.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983, sep). Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-13*(5), 834–846.

Bellman, R. (1957). *Dynamic programming* (R. Corporation, Ed.). Princeton, N.J. : London: Princeton, N.J. : Princeton U.P.

Berke, J. D. (2018, jun). *What does dopamine mean?* (Vol. 21) (No. 6). Nature Publishing Group.

Bush, P. C., & Sejnowski, T. J. (1995, oct). Models of Cortical Networks. In M. J. Gutnick & I. Mody (Eds.), *The cortical neuron* (pp. 174–190). Oxford University Press.

Cannon, W. B. (1927, dec). The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *The American Journal of Psychology*, *39*(1/4), 106.

Cannon, W. B. (1929, dec). Bodily Changes in Pain, Hunger, Fear, and Rage. *Archives of Internal Medicine*, *44*(6), 908.

Cannon, W. B. (1931). *Again the James-Lange and the thalamic theories of emotion.* (Vol. 38).

Cannon, W. B., Lewis, J. T., & Britton, S. W. (1927, sep). The Dispensability of the Sympathetic Division of the Autonomic Nervous System. *The Boston Medical and Surgical Journal*, *197*(13), 514–515.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014, jun). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Emnlp 2014 - 2014 conference on empirical methods in natural language processing, proceedings of the conference* (pp. 1724–1734).

Chollet, F. (2015). Keras: The Python Deep Learning library. *Keras.Io*.

Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological review*, *120*(1), 190–229.

Collins, A. G. E., & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological review*, *121*(3), 337–66.

Connors, B. W., Landisman, C. E., & Reid, R. (1998, dec). Thalamus: Organization and Function (Vol. 1), Experimental and Clinical Aspects (Vol. 2). *Trends in Neurosciences*, *21*(12), 539–540.

Cools, R., Nakamura, K., & Daw, N. D. (2011, jan). Serotonin and dopamine: unifying affective, activational, and decision functions. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *36*(1), 98–113.

Dalgleish, T. (2004, jul). The emotional brain. *Nature Reviews Neuroscience*, *5*(7), 583–9.

Damasio, A. (1997). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Vintage.

Damasio, A. (2008). *Descartes' error: Emotion, reason and the human brain*. Vintage.

Darwin, C. (1859). *On the Origin of Species* (J. Murray, Ed.). London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals* (J. Murray, Ed.). London: John Murray.

Davidson, R. J. (2003, sep). Affective neuroscience and psychophysiology: Toward a synthesis. In *Psychophysiology* (Vol. 40, pp. 655–665). Society for Psychophysiological Research.

Davidson, R. J., Scherer, K. R., & Goldsmith, H. H. (2003). *Handbook of affective sciences*. Oxford ; New York: Oxford University Press.

Daw, N. D., Kakade, S., & Dayan, P. (2002, jun). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*(4-6), 603–616.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Massachusetts Institute of Technology Press.

den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., … Cools, R. (2013, nov). Dissociable effects of dopamine and serotonin on reversal learning. *Neuron*, *80*(4), 1090–100.

Descartes, R. (1637). Discourse on the Method. *Philosophy*, *1*, xii, 249.

Descartes, R. (1649). *The Passions of the Soul*. Hackett Publishing.

Di Pietro, N. C., & Seamans, J. K. (2011, jun). Dopamine and serotonin interactively modulate prefrontal cortex neurons in vitro. *Biological psychiatry*, *69*(12), 1204–11.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A., & Suarez, H. H. (1995, aug). Recurrent excitation in neocortical circuits. *Science*, *269*(5226), 981–985.

Ekman, P. (1973). *Darwin and facial expression, a century of research in review* (Vol. 23) (No. 4). Academic Press.

Ekman, P. (1992). Are There Basic Emotions? *Psychological Review*, *99*(3), 550–553.

Eliasmith, C., & Anderson, C. H. (2003). *Neural Engineering: Computation Representation and Dyamics in Neurobiological Systems* (Vol. 19) (No. 5). MIT Press.

Fellous, J.-m. (1999). The Neuromodulatory Basis of Emotion. *The Neuroscientist*, *5*, 283–294.

Fellous, J.-M., & Arbib, M. A. (2005a). *Who needs emotions?: The Brain Meets the Robot*. Oxford University Press.

Fellous, J.-M., & Arbib, M. A. (2005b, apr). "Edison" and "Russell". In J.-M. Fellous & M. A. Arbib (Eds.), *Who needs emotions? the brain meets the robot* (pp. 3–8). Oxford University Press.

Fellous, J.-m., & LeDoux, J. E. (2005, mar). Toward Basic Principles for Emotional Processing. In J.-M. Fellous & M. A. Arbib (Eds.), *Who needs emotions?: The brain meets the robot* (pp. 79–116). Oxford University Press.

Fontaine, J. J. R., Scherer, K. R., & Soriano, C. (2013). *Components of Emotional Meaning* (J. R. J. Fontaine, K. R. Scherer, & C. Soriano, Eds.). Oxford: Oxford University Press.

Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A. G., Parizeau, M., & Gagné, C. (2012, jul). DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research*, *13*(1), 2171–2175.

Frank, M. J. (2005, jan). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of cognitive neuroscience*, *17*(1), 51–72.

Frank, M. J., & Claus, E. D. (2006, apr). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological review*, *113*(2), 300–26.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001, jun). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive*, *1*(2), 137–160.

Franklin, S., Madl, T., D'Mello, S., & Snaider, J. (2014, mar). LIDA: A Systems-level

Architecture for Cognition, Emotion, and Learning. *IEEE Transactions on Autonomous Mental Development*, *6*(1), 19–41.

Freitas, J. D., Gudwin, R., & Queiroz, J. (2005). Emotion in artificial intelligence and artificial life research: Facing problems. *IVA*.

Frijda, N. H. (1986). *The Emotions*.

Frijda, N. H., & Parrott, W. G. (2011, sep). Basic Emotions or Ur-Emotions? *Emotion Review*, *3*(4), 406–415.

Girard, B., Cuzin, V., Guillot, A., Gurney, K. N., & Prescott, T. J. (2003, dec). A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, *2*(2), 179–200.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Journal of machine learning research* (Vol. 15, pp. 315–323).

Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*, *5*(1), 1–48.

Goertzel, B. (2015). *Artificial General Intelligence.*

Goleman, D. (1995). Emotional intelligence. *Personality and Individual Differences*, *9*, 1091–1100.

Grillner, S., Robertson, B., & Stephenson-Jones, M. (2013). The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *The Journal of Physiology*.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001a, may). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, *84*(6), 401–410.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001b, may). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biological cybernetics*, *84*(6), 411–423.

Gurney, K., Prescott, T. J., Wickens, J. R., & Redgrave, P. (2004). Computational models of the basal ganglia: from robots to membranes. *Trends in Neurosciences*, *27*(8), 453–459.

Gutnick, M. J., & Mody, I. (2012). *The Cortical Neuron* (M. J. Gutnick & I. Mody, Eds.). Oxford University Press.

Harlow, J. (1848). Passage of an Iron Rod through the Head. *The Boston Medical and Surgical Journal*, *39*(20), 389–393.

Harlow, J. M. (1868). Recovery from the passage of an iron bar through the head. *Publications of the Massachusetts Medical Society*, *2*, 327–347.

Hawkins, J., & Blakeslee, S. (2007). *On intelligence*. Times Books.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, *139*(1), 105–118.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews*, *34*(5), 701–720.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2011). Toward an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. In *Modelling natural action selection* (pp. 239–263).

Hewitt, C., Bishop, P., & Steiger, R. (1973). A Universal Modular ACTOR Formalism for Artificial Intelligence. In *Proceeding ijcai'73 proceedings of the 3rd international joint conference on artificial intelligence* (pp. 235–245). Morgan Kaufmann Publishers Inc.

Hochreiter, S., & Schmidhuber, J. (1997, nov). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Holovaty, A., & Willson, S. (2017). *Django.* Django Software Fundation.

Houk, J. C. (2011). Action selection and refinement in subcortical loops through basal ganglia and cerebellum. In *Modelling natural action selection.*

Ivanovic, M., Budimac, Z., Radovanović, M., Kurbalija, V., Dai, W., Bădică, C., … Mitrović, D. (2015). Emotional agents — state of the art and applications. *Computer Science and Information Systems*, *12*(4), 1121–1148.

Izard, C. E. (1993, jan). Four systems for emotion activation: cognitive and noncognitive processes. *Psychological review*, *100*(1), 68–90.

James, W. (1884). What is an Emotion? *Mind*, *9*, 188–205.

Joel, D., Niv, Y., & Ruppin, E. (2002, jun). *Actor-critic models of the basal ganglia: new anatomical and computational perspectives* (Vol. 15) (No. 4-6). Pergamon.

Jones, E. G. (1985). *The thalamus.* New York: New York.

Karten, H. J., & Shimizu, T. (1991). Are Visual Hierarchies in the Brains of the Beholders? Constancy and Variability in the Visual System of Birds and Mammals. In *The changing visual system* (pp. 51–59). Boston, MA: Springer US.

Kingma, D. P., & Ba, J. L. (2015, dec). Adam: A Method for Stochastic Optimization. In *3rd international conference on learning representations, iclr 2015 - conference track proceedings.* International Conference on Learning Representations, ICLR.

Kluver, H., & Bucy, P. C. (1937). "Psychic blindness" and other symptoms following bilateral temporal lobectomy in rhesus monkeys. *American Journal of Physiology*, *19*, 352–353.

Kreibig, S. D. (2010). *Autonomic nervous system activity in emotion: A review.*

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987, sep). SOAR: An architecture for general intelligence. *Artificial Intelligence*, *33*(1), 1–64.

Lange, C. G. (1885). No Title. *The Emotions*, 33–90.

Lazarus, R. S. (1966). *Psychological stress and the coping process.* New York: New York : McGraw-Hill, (1966).

LeDoux, J. E. (1986). Sensory systems and emotion: A model of affective processing. *Integrative Psychiatry*, *4*(4), 237–243.

LeDoux, J. E. (1990). Information flow from sensation to emotion plasticity in the neural computation of stimulus values. In *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 3–52). Bradford Books (MIT Press).

LeDoux, J. E. (1992, apr). Brain mechanisms of emotion and emotional learning. *Current Opinion in Neurobiology*, *2*(4), 191–197.

LeDoux, J. E. (1995, jan). Emotion: clues from the brain. *Annual Review of Psychology*, *46*(1), 209–235.

LeDoux, J. E. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* Phoenix.

LeDoux, J. E. (2003). *The emotional brain, fear, and the amygdala* (Vol. 23).

LeDoux, J. E. (2012, feb). Rethinking the emotional brain. *Neuron*, *73*(4), 653–76.

LeDoux, J. E. (2014, oct). Low roads and higher order thoughts in emotion. *Cortex*, *59*, 214–215.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., … Wierstra, D. (2016, sep). Continuous control with deep reinforcement learning. In *4th international conference on learning representations, iclr 2016 - conference track proceedings.*

Litt, A., Eliasmith, C., & Thagard, P. (2006). Why losses loom larger than gains: modeling neural mechanisms of cognitive-affective interaction. *Proceedings of the Twenty-eighth Annual Meeting of the Cognitive Science Society*, 495–500.

Litt, A., Eliasmith, C., & Thagard, P. (2007). A large-scale neurocomputational model of emotional decision making. *BMC Neuroscience*, *8*(Suppl 2), P14.

Litt, A., Eliasmith, C., & Thagard, P. (2008). Neural affective decision theory: Choices, brains, and emotions. *Cognitive Systems Research*, *9*(4), 252–273.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the 30 th International Conference on Machine Learning*, *30*(1), 6.

MacLean, P. D. (1949, nov). Psychosomatic Disease and the "Visceral Brain". *Psychosomatic Medicine*, *11*(6), 338–353.

MacLean, P. D. (1952, nov). Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroencephalography and Clinical Neurophysiology*, *4*(4), 407–18.

MacLean, P. D. (1970). The triune brain, emotion and scientific bias. In F. Schmitt (Ed.), *The neurosciences: Second study program* (pp. 336–349). New York: The Rockefeller University Press.

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective and Behavioral Neuroscience*.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In K. R. Scherer, T. Bänziger, & E. Roesch (Eds.), *A blueprint for affective computing: A sourcebook and manual.* Oxford University Press.

Martinez-Miranda, J., & Aldea, A. (2005, mar). Emotions in human and artificial intelligence. *Computers in Human Behavior*, *21*(2), 323–341.

McClure, S. M., York, M. K., & Montague, P. R. (2004, jun). The Neural Substrates of Reward Processing in Humans: The Modern Role of fMRI. *The Neuroscientist*, *10*(3), 260–268.

McDougall, W. (1921). *An Introduction to Social Psychology*. J.W. Luce & Company.

Mehrabian, A. (1980). *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies.* (G. &. H. Oelgeschlager, Ed.). Cambridge, Mass.

Mehrabian, A. (1996). Pleasure-Arousal-Dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261–292.

Miller, B. L., & Goldberg, D. E. (1995). Genetic algorithms, tournament selection, and the effects of noise. *Complex Systems*, *9*, 193–212.

Mnih, V., Badia, A. P., Mirza, L., Graves, A., Harley, T., Lillicrap, T. P., … Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. *33rd International Conference on Machine Learning, ICML 2016*, *4*, 2850–2869.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013, dec). *Playing Atari with Deep Reinforcement Learning.*

Nagel, T. (1974, oct). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435.

Newell, A. (1973, jan). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. *Visual Information Processing*, 283–308.

Niv, Y. (2009, jun). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007, apr). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–20.

Northcutt, G. R., & Kaas, J. H. (1995, sep). The emergence and evolution of mammalian neocortex. *Trends in Neurosciences*, *18*(9), 373–379.

Oatley, K., & Johnson-laird, P. N. (1987, mar). Towards a Cognitive Theory of Emotions. *Cognition & Emotion*, *1*(1), 29–50.

O'Doherty, J. P., Lee, S. W., & McNamee, D. (2015). *The structure of reinforcement-learning* *mechanisms in the human brain* (Vol. 1).

O'Reilly, R. C., & Frank, M. J. (2006, feb). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural computation*, *18*(2), 283–328.

O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (2007). PVLV: The Primary Value and Learned Value Pavlovian Learning Algorithm. *Behavioral Neuroscience*, *121*(1), 31–49.

O'Reilly, R. C., & Munakata, Y. (2013). *Computational Cognitive Neuroscience* (No. 2001). Wiki Book.

O'Reilly, R. C., Wyatte, D., & Rohrlich, J. (2014, jul). *Learning Through Time in the Thalamocortical Loops.*

Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions* (Vol. 18). Cambridge: Cambridge University Press.

Ortony, A., Norman, D. A., & Revelle, W. (2005). Affect and Proto-Affect in Effective Functioning. In *Who needs emotions?: The brain meets the robot* (pp. 173–202). Oxford University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*, 315–331.

Osgood, C. E. C. E. (1957). *The measurement of meaning* (G. J. Suci & P. H. Tannenbaum, Eds.). Urbana [Ill.]: Urbana Ill.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford University Press.

Panksepp, J. (2005, mar). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80.

Panksepp, J., Mir, D., Delgado, J. M. R., & Edwards, S. (1982, sep). Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, *5*(03), 407.

Papez, J. W. (1937, jan). A proposed mechanism of emotion. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *7*(1), 103–12.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* Oxford University Press: Humphrey Milford.

Picard, R. W. (2000). *Affective Computing.* MIT Press.

Plutchik, R. (1958, mar). Section of psychology: outlines of a new theory of emotion*. *Transactions of the New York Academy of Sciences*, *20*(5 Series II), 394–403.

Plutchik, R. (1962). *The Emotions: Facts and Theories, and a New Model.*

Plutchik, R., & Conte, H. (1997). *Circumplex models of personality and emotions.*

Potter, M. A., & De Jong, K. A. (2000, mar). Cooperative coevolution: an architecture for evolving coadapted subcomponents. *Evolutionary computation*, *8*(1), 1–29.

Prescott, T. J., Gurney, K., Montes-Gonzalez, F., Humphries, M., & Redgrave, P. (2002). *The robot basal ganglia: Action selection by an embedded model of the basal ganglia* (Vol. 52).

Prescott, T. J., Montes González, F. M., Gurney, K., Humphries, M. D., & Redgrave, P. (2006, jan). A robot model of the basal ganglia: Behavior and intrinsic processing. *Neural Networks*, *19*(1), 31–61.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023.

Reisenzein, R. (2006, nov). Arnold's theory of emotion in historical perspective. *Cognition & Emotion*, *20*(7), 920–951.

Rescorla, R. (2008, mar). Rescorla-Wagner model. *Scholarpedia*, *3*(3), 2237.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning ii current research and theory* (Vol. 21, pp. 64–99).

Rohrer, B. (2010). What we might look for in an AGI benchmark. *AGI Conference*.

Rolls, E. T. (2007). *Emotion Explained*. Oxford University Press.

Rolls, E. T. (2013). *Emotion and Decision-making Explained*.

Romanski, L. M., & LeDoux, J. E. (1992). Equipotentiality of thalamo-amygdala and thalamo-cortico-amygdala circuits in auditory fear conditioning. *Journal of Neuroscience*, *12*(11), 4501–4509.

Russell, J. a. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, *76*(5), 805–819.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Sander, D., Grandjean, D., & Scherer, K. R. (2005, may). A systems approach to appraisal mechanisms in emotion. *Neural networks : the official journal of the International Neural Network Society*, *18*(4), 317–52.

Saper, C. B. (2009, jan). Hypothalamus. *Scholarpedia*, *4*(1), 2791.

Saper, C. B. (2012). Hypothalamus. In J. Mai & G. Paxinos (Eds.), *The human nervous system* (pp. 548–583). Academic Press.

Saper, C. B., & Lowell, B. B. (2014, dec). The hypothalamus. *Current Biology*, *24*(23), R1111–R1116.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, *69*, 379–399.

Scherer, K. R. (1984). On the Nature and Function of Emotion: A Component Process Approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Psychology Press.

Scherer, K. R. (2005, dec). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

Scherer, K. R. (2009, dec). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1535), 3459–74.

Scherer, K. R., Bänziger, T., & Roesch, E. (2010). *A Blueprint for Affective Computing: A Sourcebook and Manual*. Oxford University Press.

Scherer, K. R., & Ekman, P. (1984). *Approaches to emotion*. L. Erlbaum Associates.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research. Series in affective science*.

Scheutz, M. (2004). Useful roles of emotions in artificial agents: A case study from artificial life. In *Proceedings of the 19th national conference on artificial intelligence* (pp. 42–47). AAAI Press.

Scheutz, M. (2012). The affect dilemma for artificial agents: Should we develop affective artificial agents? *IEEE Transactions on Affective Computing*, *3*(4), 424–433.

Schmajuk, N. (2008, mar). Classical conditioning. *Scholarpedia*, *3*(3), 2316.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017, jul). *Proximal Policy Optimization Algorithms*.

Schultz, W. (2000, dec). Multiple reward signals in the brain. *Nature reviews. Neuroscience*, *1*(3), 199–207.

Seamans, J. (2007, jun). Dopamine anatomy. *Scholarpedia*, *2*(6), 3737.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Shanahan, M. (2010). *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds*. Oxford University Press.

Sherman, S. (2006, sep). Thalamus. *Scholarpedia*, *1*(9), 1583.

Sherman, S. M. (2016, mar). Thalamus plays a central role in ongoing cortical functioning. *Nature Neuroscience*, *16*(4), 533–541.

Sherman, S. M., & Guillery, R. W. (2002, dec). The role of the thalamus in the flow of information to the cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *357*(1428), 1695–708.

Sherman, S. M., & Guillery, R. W. (2005). *Exploring the Thalamus and Its Role in Cortical Function*.

Sloman, A. (1999). How Many Separately Evolved Emotional Beasties Live Within Us? *Emotions in humans and artifacts*.

Sloman, A. (2001). Beyond shallow models of emotion. *Cognitive Processing: International Quarterly of…*, *2*, 177–198.

Snaider, J., McCall, R., & Franklin, S. (2011). The LIDA framework as a general tool for AGI. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 6830 LNAI, pp. 133–142).

Stent, G. S. (1973). A Physiological Mechanism for Hebb's Postulate of Learning. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(4), 997–1001.

Suri, R., & Schultz, W. (1999, jul). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience*, *91*(3), 871–890.

Sutton, R. S. (1984). *Temporal Credit Assignment in Reinforcement Learning*.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: an Introduction*.

Swanson, L. W. (2000, dec). Cerebral hemisphere regulation of motivated behavior. *Brain Research*, *886*(1-2), 113–164.

Thagard, P. (1989, sep). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(03), 435–502.

Thagard, P. (2002). *Coherence in Thought and Action*. Place of publication not identified MIT Press, 2002.

Thagard, P., & Aubie, B. (2008, sep). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition*, *17*(3), 811–834.

Thorndike, E. L. (1911). *Animal intelligence*. New York: The Macmillan company.

Toga, A. W. (2015). *Brain mapping : an encyclopedic reference* (Vol. 1-3).

Treue, S. (2001, may). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, *24*(5), 295–300.

Turian, J., Bergstra, J., & Bengio, Y. (2009). Quadratic features and deep architectures for chunking. In *Human language technologies: Conference of the north american association of computational linguistics* (p. 245). Association for Computational Linguistics (ACL).

Verschure, P., & Eds, A. D. (2014). *Distributed Adaptive Control: Theory and Practice*. Barcelona, Spain.

Vitay, J., & Hamker, F. H. (2010). *A computational model of Basal Ganglia and its role in memory retrieval in rewarded visual memory tasks*.

Wagar, B. M., & Thagard, P. (2004, jan). Spiking Phineas Gage: a neurocomputational theory of cognitive-affective integration in decision making. *Psychological review*, *111*(1), 67–79.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015, apr). A Bayesian Model of Category-Specific Emotional Brain Responses. *PLoS computational biology*, *11*(4), e1004066.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards* (Unpublished doctoral dissertation). Cambridge, England.

Wells, F. L. (1925). Reactions to visual stimuli in affective settings. *Journal of Experimental Psychology*, *8*(1), 64–76.

Wells, F. L., Forbes, A., & Whitley, M. T. (1912, jan). On Certain Electrical Processes in the Human Body and Their Relation to Emotional Reactions. *The American Journal of Psychology*, *23*(1), 149.

Werbos, P. J. (1990). Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE*, *78*(10), 1550–1560.

Whalen, P. J., & Phelps, E. A. (2009). *The Human Amygdala*. Guilford Press.

Williams, R. J. (1992). Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Machine Learning*, *8*(3), 229–256.

Woergoetter, F., & Porr, B. (2008, mar). Reinforcement learning. *Scholarpedia*, *3*(3), 1448.

Woodworth, R. S., & Sherrington, C. S. (1904, jun). A pseudaffective reflex and its spinal path. *The Journal of Physiology*, *31*(3-4), 234–243.

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie, Volume 1*. W. Engelman.

Yates, J. R., Darna, M., Gipson, C. D., Dwoskin, L. P., & Bardo, M. T. (2015, may). Dissociable roles of dopamine and serotonin transporter function in a rat model of negative urgency. *Behavioural brain research*, *291*, 201–208.

Yilmazer-Hanke, D. M. (2012). Amygdala. In J. K. Mai & G. Paxinos (Eds.), *The human nervous system (third edition)* (Third Edit ed., pp. 759–834). San Diego: Academic Press.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

# APPENDIX A

## The **ProtoEmo** architecture in detail

As stated in Chapter 1 the goal of this thesis is to explore the role emotions play in intelligence. More specifically, the objective is to show that emotions inform the brain as to the nature of a given situation, and guide the decision-making process to increase the chances of survival for a virtual agent. However, as made abundantly clear in Chapter 2 the concept of emotion does not have a standard definition yet. Moreover, most of the theories and models focus on emotions at a higher-level where cognition already interferes (according to the model suggested by Ortony et al. (2005) and presented in Section 2.6.3, but see Alexandrov and Sams (2005), and Izard (1993) for similar ideas). Consequently, before investigating the role of emotions in the decision-making process it is important to understand the low-level mechanisms triggering emotional episodes.

**ProtoEmo** is the artificial neural architecture introduced in Chapter 5. It was implemented as a controller in a resource foraging task to explore the relevance of the '*Survival Circuits*' for virtual agents, as well as the role of the amygdala in those homeostatic mechanisms. LeDoux (2012) defined the concept of survival circuits (see Sub-Section 2.6.2 for an in-depth description of this theory), based on ideas from the '*basic emotions*' theory (presented in Sub-Section 2.6.1), as well as Darwin (1872) suggestion that emotions are a mechanism that evolved to improve the survival potential of animals. Briefly, in the animal's brain survival circuits are responsible for maintaining the body's homeostasis. As such, they are capable of detecting stimuli relevant to the survival of the individual. Upon triggering, a survival circuit sends signals to recruit different brain systems in an attempt to bring the body's state back to balance or defend against danger.

As a continuation to Section 5.1 this appendix dives deeper into the design decisions made during the development of the **ProtoEmo** architecture (shown in Figure A.1), as well as specify all the elements required for its implementation. The final portion of this appendix presents the genetic algorithm used as the learning process for the **ProtoEmo** architecture.

## A.1 From brain areas to layers of artificial neurons

Before explaining how the activation values on the output layer of **ProtoEmo** are used to build a strategy, this section first describe the basic building block used for implementing the architecture. Details about the brain areas included in **ProtoEmo**, their connectome and the layers used to model them are reported here as well.

### A.1.1 Basic building block: the artificial neuron

The basic building block used throughout the implementation of the **ProtoEmo** architecture is the artificial neuron. To compute its output activity each neural unit applies a Leaky ReLU function (introduced by Maas et al. (2013)) to the weighted sum of its inputs. The Leaky ReLU activation function used for this implementation can be described as:

$$f(x) = \begin{cases} \alpha x & \text{if } x < 0 \\ x & \text{otherwise} \end{cases} \tag{A.1}$$

Where $\alpha$ the negative slope coefficient has been set to $0.2$. There are two main reasons for using a ReLU function to compute the output activity of neurons within the **ProtoEmo** architecture. First, according to Glorot, Bordes, and Bengio (2011) ReLU is a good enough approximation for modeling cortical neurons. Indeed, Bush and Sejnowski (1995) have shown that pyramidal cells from the visual cortex of cats rarely reach their saturation regime. Furthermore, Douglas, Koch, Mahowald, Martin, and Suarez (1995) using anatomical evidences developed an electrical circuit to model the firing rate of pyramidal cells, while taking into account both recurrent and inhibitory connections to a neuron. Thanks to their electrical circuit they were able to show that the frequency at which cortical neurons fire action potentials is proportional to its input current. Second, when compared to other standard activation functions, such as the sigmoid

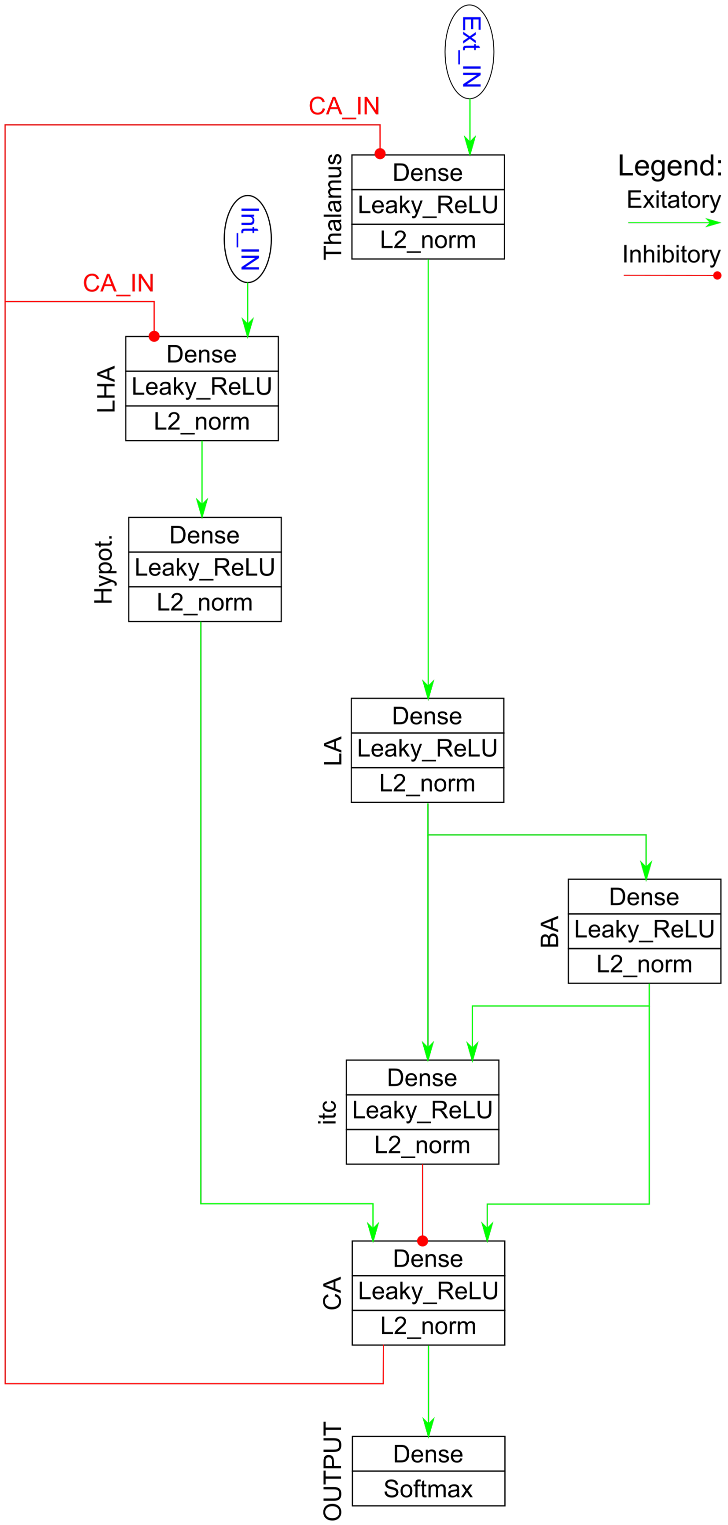


Figure A.1 *A visual representation of the layers implemented for the* ***ProtoEmo*** *controller. It should be noted that the order of appearance is the same as the sequence followed to propagate the activities of the input units through the architecture.*

function or tanh, the Leaky ReLU function takes less CPU time to compute. This might at first seems like a small time save, but it can become quite significant when hundreds of virtual agents are using hundreds of neurons for thousands of simulation cycle. Even on modern computers, using Leaky ReLU as an activation function instead of the sigmoid function, for example, could mean the difference between a simulation that complete in a matter of minutes and one that takes an hour.

Moving a level higher in the organizational hierarchy, in **ProtoEmo** neurons are arranged in fully-connected layers. Moreover, within a layer each neural unit can influence the activity of all the others via lateral inhibition. In animals, the vast majority of neurons in the neocortex are excitatory pyramidal cells (Bush & Sejnowski, 1995; Douglas et al., 1995; Gutnick & Mody, 2012). Those neurons send 80% of their connections to other excitatory cells within a $1mm$ neighborhood referred to as a '*column*'. However, according to Douglas et al. (1995) for cortical columns to be able to perform any sort of useful computation, pyramidal cells need to project to and receive activity from inhibitory neurons. A fact that Armony et al. (1995) also highlighted in their experiment (see Chapter 4 for more details). When designing their computational model of the dual pathways to the amygdala, Armony et al. (1995) realized that by introducing a graded version of the '*winner-takes-all*' mechanism, in lieu of lateral inhibition, neural units developed what they called a '*receptive field*'. In the context of Armony et al.'s (1995) experiment, it means that a neuron is strongly activated by a single frequency (called the '*best frequency*'), but will also respond weakly to adjacent frequencies. Furthermore, Douglas et al. (1995) differentiates between two types of inhibition:

> "*The feedback inhibition mediated by linear synapses acts as a multiplicative, shunting-like inhibition that changes the gain of cortical response to a given input current, whereas linear feedforward inhibition acts more like an offset that reduces the input current by a given amount.*" — (Douglas et al., 1995, p.982-983)

This reliance of a brain area on inhibition to perform its duty, echoes suggestions made about the inner workings of the amygdala. Indeed, due to the feedback and feedforward inhibitions the central nucleus of the amygdala is subjected to (from the inter-neurons and the intercalated nucleus, respectively) only novel or strongly energized stimuli can elicit a burst of activity (Whalen & Phelps, 2009). As a result, neurons within the amygdala are usually described as "*silent*" (Whalen & Phelps, 2009). This is in line with the accepted role of the amygdala in animals. As discussed at length in both Chapter 2 and Chapter 5 (and summarized later in this appendix), the amygdala is ideally placed to detect stimuli relevant to the survival of the individual and warn the rest of the brain whenever such stimulus appears. Consequently, since most inhibition in **ProtoEmo** can be considered feedback, lateral inhibition has been modeled using the $L2$-norm. For an activity vector of size $n$ the $L2$-norm is defined as:

$$|x_i| = \frac{x_i}{\sqrt{\sum_{k=0}^{n} x_k^2}} \tag{A.2}$$

As shown in Equation A.2, to determine the final activity of a single neural unit, therefore, the activation values of all the neurons in a layer are taken into account.

So far the implementational details corresponding to **ProtoEmo**'s basic building block, the artificial neuron, as well as the layers that make up the architecture have been described. Hence, the remainder of this section describes the role and the connectome of the different brain areas modeled (see also Figure A.1 for an overview of the connections implemented).

### A.1.2 The thalamus

The first brain region to be included in the **ProtoEmo** architecture is the thalamus. In animals the accepted role of the thalamus is to redirect information coming from the senses to the corresponding brain systems. Indeed, except for the olfactory system, all sensory data projects in a topographically organized manner onto the thalamus, which then dispatches the filtered information to the targeted cortical and sub-cortical areas (J. R. Anderson et al., 2004; Armony et al., 1995; Connors, Landisman, & Reid, 1998; Jones, 1985; O'Reilly & Munakata, 2013; O'Reilly, Wyatte, & Rohrlich, 2014; S. Sherman, 2006; S. M. Sherman, 2016; S. M. Sherman & Guillery, 2002, 2005; Toga, 2015). As a result, the thalamus was the focus of many theories and models of emotions (see Section 2.6 and Chapter 4 for more details). Within the **ProtoEmo** architecture, the thalamus is represented by a single layer of 20 neurons. It receives inhibitory connections from the central nucleus of the amygdala, as well as the following external inputs describing the state of the environment:

- the distance to the closest obstacle,
- the distance to the closest source,
- the distance to the closest robot,
- and finally, the maximum action tendency sampled over all neighboring agents, if any. Otherwise, this input defaults to zero.

It should be noted that only the distances between the current agent and the different objects are necessary to build a complete strategy, since the agent does not decide on the exact direction and speed of its movement, but only influences the computation of its velocity, as has been described in Section 5.2. Given that the sensory modality of the thalamic inputs to the **ProtoEmo** controller are similar in nature, a single layer is enough to model the thalamus. Finally, given the lack of neocortical areas needed for further refinement of the external sensory information, the thalamus only projects to the lateral nucleus of the amygdala.

### A.1.3 The hypothalamus

After focusing on the (external) state of the environment, the next brain region looks inward to ensure the survival of the individual and the species. The hypothalamus has been described as the control center for the Autonomic Nervous System (ANS). In animals this system is in charge of maintaining the body's homeostasis or overcome stressors (a process also referred to as '*allostasis*') (Dalgleish, 2004; J.-m. Fellous, 1999; Kreibig, 2010; Saper, 2009, 2012; Saper & Lowell, 2014; Swanson, 2000). To perform its duty, the hypothalamus receives inputs from a wide variety of sensory regions describing the state of the individual. The main recipient of those sensory inputs is the lateral hypothalamic area (LHA) (Saper, 2012; Saper & Lowell, 2014). Moreover, the hypothalamus can be seen as an aggregation of many '*centers*', each responsible for a different regulatory function. Thus, whenever a center detects a discrepancy it activates an autonomic, endocrine, or behavioral response to bring the internal milieu back to balance (Saper, 2012; Saper & Lowell, 2014; Swanson, 2000). As such, the hypothalamus is not really a '*control center*', since it only influences the individual's motivation to perform certain actions (J.-m. Fellous, 1999; Saper, 2012; Saper & Lowell, 2014). Among the centers involved in the different homeostatic and allostatic functions, the medial hypothalamus plays an important role in energy regulation, and in defensive, aggressive and reproductive behaviors. The preoptic area is implicated not only in reproductive behavior as well, but also in thermoregulation, and in arousal (Saper, 2009, 2012; Saper & Lowell, 2014; Swanson, 2000). Therefore, within the **ProtoEmo** architecture the hypothalamus has been split into two layers. The LHA, with a size of 10 neurons, receives the sensory inputs describing the agent's internal state. The second layer, represents the union of the medial hypothalamus and preoptic area. It contains five neural units, and has simply been labeled as Hypothalamus (Hypot) in Figure A.1. The features used to describe the internal state of virtual agents in the resource foraging task are:

- the current energy level of the robot,

- the current value of the action tendency,
- and a computed value defined as:

$$\begin{aligned} Danger = {} & (2 \times obs\_thres - \min(2 \times obs\_thres, obs\_dist)) \\ & \times (3 \times fight\_thres - \min(3 \times fight\_thres, rbt\_dist)) \end{aligned} \tag{A.3}$$

Where $obs\_thres = 0.5$ is an arbitrary threshold warning the robot of an incoming obstacle, and $fight\_thres = 0.3$ is the distance between two agents under which they are considered to be in a battle situation.

In addition to taking care of the internal balance of an individual, it has been suggested that the hypothalamus also plays a major role in the association of drives and reward (J.-m. Fellous, 1999). An idea that echoes the role attributed to the LHA by O'Reilly et al. (2007) and Hazy et al. (2010) in their PVLV architecture (see Sub-Sub-Section 3.3.2 for a detailed explanation on this model). Given its position and influence over the ANS, neuroendocrine, and behavioral systems, the hypothalamus has been considered by some as the locus for emotional expression (J.-m. Fellous, 1999).

Aside from receiving inputs from internal sensory devices, the LHA is also the target of inhibitory projections coming from the central nucleus of the amygdala (CA). Activation values from the LHA are directly sent to the Hypot layer, which then connects to the CA. Hence, creating a feedback loop in which the amygdala can influence the sensory data entering the hypothalamus.

### A.1.4 The amygdala

The amygdala is the last brain region to be included in the **ProtoEmo** architecture. This area has grabbed the attention of many researchers in the affective science community, and as a target in fear conditioning experiments. In anatomic theories of emotions (discussed in Section 2.6) the amygdala is usually construed as simply processing emotions. However, based on observations from fear conditioning studies researchers have been able to map out its structure, and have suggested its role to be similar to an alarm system (Dalgleish, 2004; J.-M. Fellous & Arbib, 2005a; O'Reilly & Munakata, 2013; Sloman, 1999; Whalen & Phelps, 2009; Yilmazer-Hanke, 2012). Indeed, as explored in more details in Chapter 5, the amygdala is ideally placed to detect stimuli that are relevant to the survival of the individual and the species. Detection of a significant stimulus leads to the activation of the amygdala's central nucleus (CA), which sends action potentials to a wide array of cortical and sub-cortical areas, including neuromodulatory systems (dopamine and serotonin). The release of neuromodulators in the brain alter cognitive processing, and guide the emotional and behavioral responses of the individual (Whalen & Phelps, 2009).

Within the **ProtoEmo** architecture the amygdala has been implemented as four separate layers: *1*) the lateral nucleus (LA, size 5), which is considered to be the sensory interface to the amygdala, since it receives connections from the thalamus; *2*) the basal nucleus (BA, size 10), which is an aggregation of the basal and accessory basal nuclei found in the animal's brain; *3*) the intercalated nucleus (Itc, size 10); *4*) and finally, the central nucleus (CA, size 10), which is also considered to be the output of the amygdala. At an abstract level, the amygdala's structure has been compared to a funnel. Sensory information arriving from the LA has to travel via the BA to reach the CA. Each traversed module further reducing the dimension of its input data and extracting higher-level concepts. Both the LA and BA also target the Itc, which in turn send inhibitory projections to the CA. According to Whalen and Phelps (2009), and O'Reilly and Munakata (2013), this feedforward inhibition implements a habituation mechanism. As a result, only strongly activating stimuli can lead the amygdala to send action potentials. Hence, earning the amygdala the label of '*silent*' (Whalen & Phelps, 2009).

As depicted in Figure A.1, in terms of extrinsic connections the amygdala receives sensory data describing the environment's state from the thalamus, which projects onto the LA. Information related to the state of the individual is relayed directly to the CA via the Hypot layer. Finally, as mentioned in Section 5.1 the CA sends inhibitory projections that target both the thalamus and LHA. Via those connections the amygdala is able to '*filter*' the sensory data arriving to the CA. Hence, implementing two primary attention mechanisms. The final excitatory synapse links the CA to the output layer.

### A.1.5 Outputs

The output of the **ProtoEmo** architecture is the only layer that differs from the rest. It contains four neural units, whose activation values result from the application of the softmax function over the weighted sum of their respective inputs. The standard softmax function, for a vector $\vec{x}$

of size $n$, is defined by the following equation:

$$f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \text{ for } i = 1, \ldots, n \tag{A.4}$$

Usually the softmax function is used on a layer when its output values are meant to represent probabilities, since a property of the softmax is that the sum of the elements of the resulting vector is always equal to 1. Though, in the context of **PrimEmo** as a controller there were only three reasons for choosing the softmax activation function. The first one is that the activation values of the output layer had to fall within the $[0,1]$ range. To achieve this effect, a standard sigmoid function could have been used as well. However, the sigmoid function has a steeper curve between its boundaries compared to the softmax function, which means that activation values tend to get pushed towards the extrema. This is something that is also present, but to a lesser degree with the softmax function. Lastly, using the softmax function also introduces lateral inhibition to the output layer. However, the significance of this lateral inhibition is different from the one specified for all the other layers. Indeed, for the output layer it is interpreted as a modeling of the fact that when an animal's well-being is in danger, other needs, such as eating or drinking, are actively suppressed to free up cognitive resources and efficiently deal with the aversive stimuli (LeDoux, 2012).

As explained in Section 5.2, the different controllers implemented for the resource foraging task can only influence how the controlled robots move, as well as if they stand their ground or flee in a conflict. **ProtoEmo** is no exception. Therefore, the activity of each neuron in the output layer provides the value for a different coefficient. The activation rate of the first neuron is directly assigned as the action tendency of the virtual agent. In a conflict situation, similar to Scheutz's emotional agents (described in Sub-Section 5.2.2), a robot controlled by the **ProtoEmo** architecture retrieves the action tendencies from its neighboring peers. If the action tendency of the current robot is the highest, it fights. Otherwise, the agent runs away. The other three neurons all provide values that affect the robot's movement. As stated in Sub-Section 5.2.1 for any agent the direction and speed of movement are computed using Equation 5.1. Three coefficients are part of this computation, and reflect the agent's current needs and goals. Borrowing from Scheutz's (2004) design the fighting tendency ($g_a$) is defined as:

$$g_a = 100 \times output\_activity_2 - 50 \tag{A.5}$$

Where $output\_activity_2$ is the activation value of the second neuron from the output layer. Similarly, both the tendency to avoid obstacles ($g_t$) and the forage tendency ($g_r$) are computed based on the activity of the third ($output\_activity_3$) and fourth neurons ($output\_activity_4$), respectively, as follows:

$$g_t = -20 \times output\_activity_3 \tag{A.6}$$

$$g_r = 20 \times output\_activity_4 \tag{A.7}$$

## A.2 Evolving ProtoEmo for optimal survival

Inspired by Darwin's (1872) theory dictating that emotions are a mechanism that evolved to improve the survival potential of individual and the species, a simple genetic algorithm (GA) was used for training the **ProtoEmo** controller. This draws a nice parallel between animals evolving emotions and now virtual agents evolving emotions of their own. Aside from satisfactory parallels, a GA was also chosen for two main reasons. First, due to the fact that the **ProtoEmo** architecture has two recurrent loops and does not reach stability at a fixed point in time (as explained in Sub-Section 5.2.3), it would be unwieldy and memory intensive to use backpropagation through time (Lillicrap et al., 2016; Werbos, 1990). Second, even though other meta-learning algorithms, such as simulated annealing, could be used as well, GAs have the advantage of being highly parallelizable by nature. The Distributed Evolutionary Algorithms in Python (DEAP, written by Fortin, De Rainville, Gardner, Parizeau, and Gagné (2012)) library was used for the actual implementation of said genetic algorithm. Besides streamlining the evolutionary process, it also made splitting the workload over different computing nodes much easier.

The remainder of this section first defines the individuals and population used for the genetic algorithm. The next part of this section, goes over the evaluation protocol and its associated fitness function. Finally, the operators selecting and mating individuals from each population to create the next generation are explored in details.

### A.2.1 Populations of synaptic strengths

The only means by which it is possible to teach the **ProtoEmo** controller the optimal strategy to promote the survival of the individual and the species within the environment of the resource foraging task, is to update the strengths of the synapses connecting layers together. Furthermore, since the actual implementation of the **ProtoEmo** controller uses the Keras framework (Chollet, 2015), backed by the Tensorflow library (Abadi et al., 2015), it means that the connection weights for a given layer are stored in a matrix. Consequently, the chromosomes used for the learning process take the shape of matrices implemented as lists of lists. Thus, weight matrices can be represented as:

$$\mathbf{W} = \begin{pmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,j} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i,0} & w_{i,1} & \cdots & w_{i,j} \end{pmatrix} \tag{A.8}$$

Where $i$ the number of rows represent the size of the current layer, and $j$ is the number of units projecting to the current layer. At the beginning of the training phase, each individual is independently initialized with random values sampled from a Gaussian distribution of mean $\mu = 0.5$ and a standard deviation $\sigma = 0.5$.

Rather than evolving the synaptic strengths of the whole architecture at once, the principles of cooperative coevolution (Potter & De Jong, 2000) have been integrated within the training process. In nature, the concept of coevolution is defined by two or more species reciprocally influencing each other's evolution through natural selection. Within the context of genetic algorithms, and especially in DEAP, cooperative coevolution is achieved by first defining multiple populations. Then, to evaluate the fitness of an individual from a given population it is aggregated with the representatives of other populations to form a solution to the task. Here a species' representative is construed to be the best individual found so far. Once all individuals from a species have been evaluated, selected, and mated, the algorithm moves on to the next species. A generation, therefore, consists in going through all the populations in order, evaluating and mating their individuals. When applied to the connection strengths of a neural network, like the **ProtoEmo** controller, coevolution forces each layer to find the most efficient way to use the best data that its input layers can provide. Consequently, instead of defining a single population containing the weights corresponding to all the layers, the current implementation uses one population per layer. As a result, the implemented GA simultaneously evaluates and updates eight species, comprised of 20 individuals each.

### A.2.2 Evaluating individuals

By definition, in computer science a genetic algorithm is a meta-heuristic, whose purpose is to find solutions to search and optimization problems. To find the optimal solution to a problem GAs evaluate populations of individuals and assign a '*fitness*' value to each of them. The fitness usually is a single value representing how good of a solution a particular individual is. In the

DEAP framework it is possible to define multiple objectives for a single GA. Hence, each individual can have one or more fitness values (one for each of the objectives). In the particular context of the resource foraging task, two objectives have been outlined. The most important goal for all virtual agents is to survive for as long as possible. Therefore, the first objective is to optimize for the number of survivors at the end of a simulation. To guide the GA in reaching this first goal, a second objective of maximizing the size of the group of robots in a simulation has been set as well. Between the two objectives maximizing the population of survivors takes precedence, and this has been defined in DEAP by setting a weight of 1 for the first objective and of 0.5 for the second one.

Within the GA implemented for the **ProtoEmo** controller's training phase, the evaluation process involved the following steps:

1. Build a neural network embodying the design described in Section A.1, and shown in Figure A.1, using placeholder weights for the connections.
2. As mentioned in the previous Sub-Section, since each layer is considered as a separate species the individual to be evaluated has to first be combined with the representatives of the other species.
3. The matrices representing the synaptic strengths are then assigned to their corresponding connections in the neural network.
4. Now that the individual is finally ready, it is used as the '*brain*' controlling all the robots for a resource foraging task. The task lasts for 2500 cycles.
5. At the end of the simulation, the number of surviving agents, as well as the maximum size the robot population attained during the simulation are assigned as fitnesses for the individual under evaluation.

To obtain the results presented in Section 5.3 the **ProtoEmo** controller has been trained over 200 generations. Keep in mind that a single generation entails evaluating all 20 individuals from all 8 populations. Therefore, in total around 32000 individuals have been evaluated.

### A.2.3 Making the next generation

Once all individuals inside a population have been evaluated, they are then selectively mated to produce the next generation. Beginning with the selection process, there are many ways to choose individuals to be the parents of the next generation. Usually, though, selection mechanisms use the individual's fitness to decide whether to pick it or not. This way the probability of an individual to be chosen for inclusion as a parent increase the higher its fitness is. In the GA implemented to train the **ProtoEmo** controller a '*tournament selection*' mechanism has been used with a tournament size of 3. The idea behind the tournament selection is quite straight forward. Given a population of individuals and a tournament size of $k$, first randomly select $k$ individuals from the population. Out of this sub-population, the individual with the highest fitness gets selected. To select multiple individuals simply repeat the process as many time as needed. Due to the simplicity of its implementation, as well as the ease with which selection pressure can be adjusted the tournament selection has been commonly used by GAs for a long time (Miller & Goldberg, 1995). Furthermore, with a tournament size of 3 it means that the resulting population will be a mix between individuals with high fitnesses, while still maintaining a good diversity by including individuals with lower fitnesses.

In GAs the standard mating process is usually split into two mechanisms: crossover and mutation. Crossover is an operation through which the genetic information of both parents are combined to form one or two children. In the GA implemented for the resource foraging task, a row-wise two-points crossover procedure has been implemented. This means that for every other rows (the crossover probability is set to 0.6) of both parents two indexes are selected at random. The data before the first index and after the second index is then swapped between the parent to produce two new children. After all parents have been combined to produce the next generation, the children are then mutated. Mutation as in nature randomly changes the value of a gene, to maintain genetic diversity, and hopefully not get trapped in a local optima. Given that individuals are matrices representing synaptic weights, the implemented GA employs a Gaussian mutation. Each row of an individual has a 0.1 probability of being considered for mutation. Next, each element in that row has a further 0.01 chance of being mutated. Finally, if an element is selected for mutation, then its value is replaced by another one sampled from a Gaussian distribution with mean $\mu = 0.5$ and standard deviation $\sigma = 0.25$.

After evaluating all the individuals in a population, selecting the fittest to be the parents for the next generation, combining those parents to produce offspring, and finally mutating those children to maintain genetic biodiversity, the genetic algorithm moves to the next species, and does it all over again for many generations.

# APPENDIX B

# The **PrimEmo** architecture in detail

After validating that LeDoux's (2012) '*Survival Circuits*' theory is applicable to virtual agents, and indeed improve their survival potential, using the **ProtoEmo** architecture (described in Section 5.1 and further detailed in Section A.1), the next part of this thesis finally explores the role of emotions in the decision-making process. The **PrimEmo** architecture results from the integration of the PVLV and basal ganglia models (both introduced in Sub-Section 3.3.2) with the **ProtoEmo** architecture. **PrimEmo** has been built to investigate whether virtual agents can acquire primary emotions using the brain structures afforded to them. Furthermore, if the concept of primary emotions can indeed emerge from those brain circuits, it is then hypothesized that emotions inform and guide the decision-making process. Thus, increasing the survival chances of the individual, as well as the species. Remember that in this thesis primary emotions have been differentiated from other patterns of brain activity by their level of arousal and valence (see Chapter 2 for more details).

This appendix first describes the inner workings of the **PrimEmo** architecture as it has been implemented to play Viking Doom (a game introduced in Section 6.3 and detailed in Section C.2). It then presents, in a second part, the genetic algorithm used to train the **PrimEmo** controller to survive in a complex and dynamic environment.

## B.1 The PrimEmo architecture in detail

Before outlining the layers, and connections between those, that are part of the implementation of **PrimEmo**, this section starts by defining the building block used throughout the architecture. The second sub-section addresses the differences in implementations between the **ProtoEmo** architecture presented in Sub-Section A.1 and the one integrated with the PVLV and basal ganglia models to make up **PrimEmo**. Lastly, this section focuses on the connections between the different models and their role in the animal's brain.

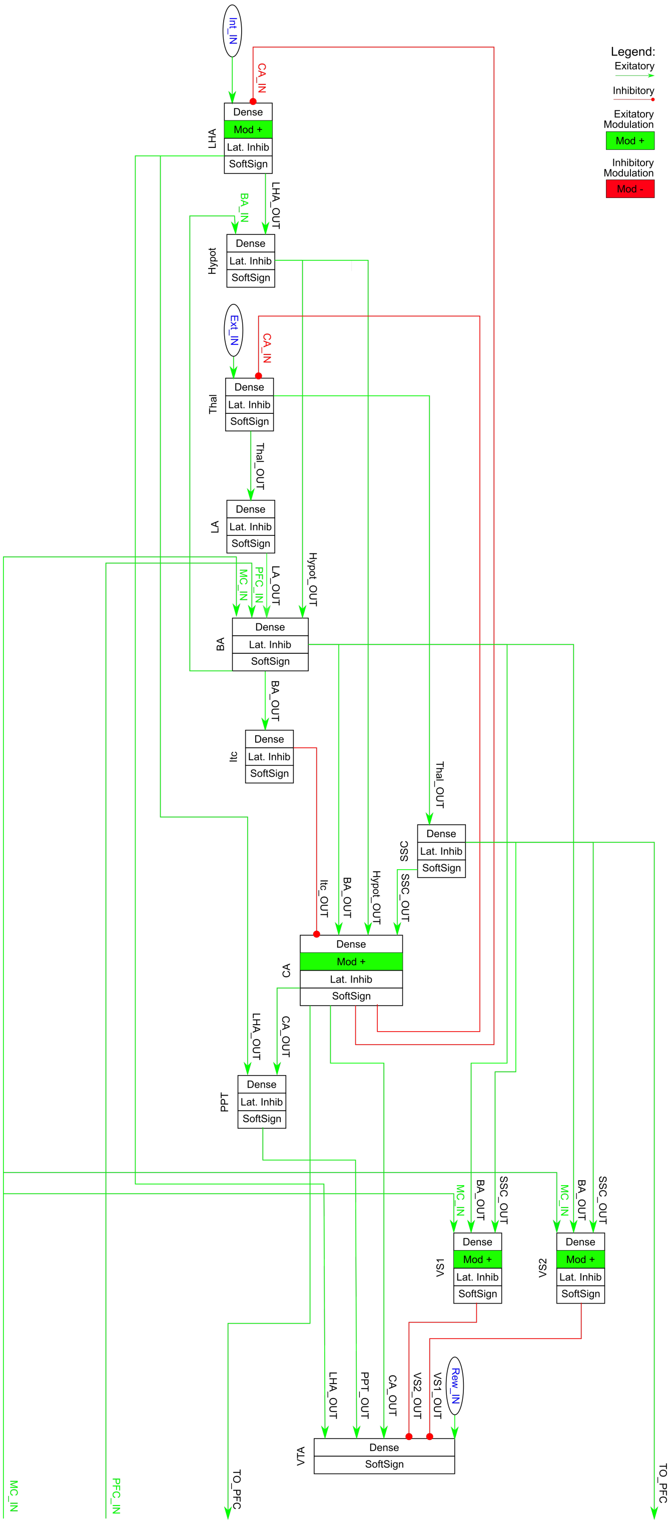
### B.1.1 Basic building block: the artificial neuron

Similar to the brain structure that it models the fundamental building block of the **PrimEmo** architecture is the neuron. The artificial neuron used in the implementation of the **PrimEmo** controller is a non-linear unit applying a '*Softsign*' activation function (Turian, Bergstra, & Bengio, 2009) to the weighted sum of its input activities. The softsign function is here defined as:

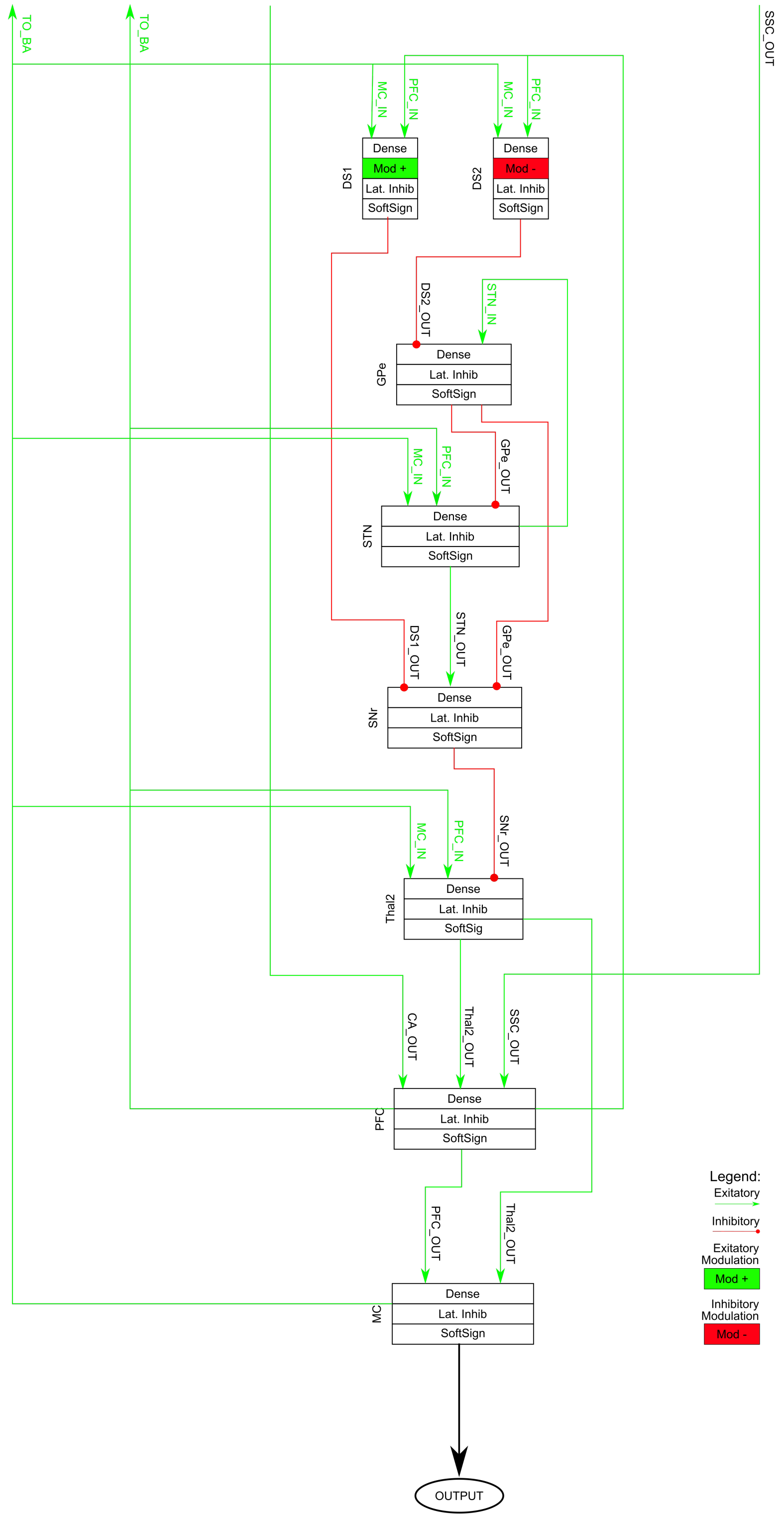
$$f(x) = \frac{x}{1+|x|} \tag{B.1}$$

On the contrary to the Leaky ReLU function used for **ProtoEmo** (defined in Sub-Section A.1.1), the output of the softsign function falls inside the range $[-1,1]$. The fact that all neurons in **PrimEmo** use a softsign activation function is an error in the implementation. Originally, the design intended for neural units to apply a ReLU function to the weighted sum of their inputs for reasons similar to those expressed in Sub-Section A.1.1. The softsign activation function should have been reserved for the neurons of the VTA layer, since they are in charge of computing the discrepancy between predicted and actual reward, as well as between predicted and actual punishment. In this context, an output value of $-1$ would represent a punishment, whereas a value of 1 would be a reward. Therefore, 0 would be considered as neutral. Although it was a mistake, softsign has an advantage over the more standard hyperbolic tangent (tanh) activation function. Compared to tanh, softsign approaches its boundaries more smoothly which means that the activation values are not pushed as much towards the limits. As a result, the neural activities are better spread across the $[-1,1]$ range.

Neural units representing the same brain area are organized in layers. Inside a layer neurons are able to influence each other's activity via lateral inhibition for reasons similar to those expressed in Sub-Section A.1.1. On the contrary to the **ProtoEmo** architecture, however, lateral inhibition is not implemented as a normalization layer, but takes the shape of laterally



(a) *The Critic portion of the* ***PrimEmo*** *architecture.*



(b) *The Actor part of the* ***PrimEmo*** *architecture.*

Figure B.1 *A visual representation of the layers implemented for the* ***PrimEmo*** *controller. It should be noted that the order of appearance is the same as the sequence followed to propagate the activities of the input units through the architecture.*

projecting synapses connecting a neuron with all other units in the layer. The synaptic strength for the lateral connections is a constant set per layer to:

$$\omega = \frac{1}{\sqrt{layer\_size}} \tag{B.2}$$

Where *layer_size* is the number of units in the layer. It should be noted that in a neuron the lateral inhibition is taken into account before applying the activation function. Additionally, wherever a link is depicted between layers in Sub-Figure B.1a the sending layer fully connects to the receiving one.

The brain areas part of the actor portion of the **PrimEmo** architecture (illustrated in Sub-Figure B.1b) are each implemented as a separate group. Inside each group layers are independent from each other. However, where group sends projections to another one, each layer in the sending group fully connects to the corresponding layer in the receiving group. Each of the five layers within a group represents one of the actions available to the hero in Viking Doom (see Sub-Section 6.3.4 for more details on what those actions are). This concept of groups of independent layers has been included in **PrimEmo**'s design to replicate the strip-like organization of both the pre-frontal cortex (PFC) and basal ganglia (BG) as explained in more details in Sub-Sub-Section 3.3.2. Moreover, according to Frank et al. (2001), O'Reilly and Munakata (2013), and Gurney et al. (2001a, 2001b) this is also a feature necessary for the action-selection mechanism implemented by the basal ganglia to perform its duty. Wherever a group of layers is depicted as projecting to a single layer in Figure B.1, the group is first flattened (the layers are merged into a single one) before fully connecting to the receiving layer.

Finally, as depicted in Figure B.1 some of the brain areas are the target of either excitatory or inhibitory modulation. Modulation within the **PrimEmo** architecture comes from the ventral tegmental area (VTA) layer, which contains a single neuron. As explained above, this neuron applies a softsign activation function to the weighted sum of the activities on its input. Since the values produced by the softsign function all fall within the $[-1,1]$ range, 1 is considered as the best reward possible, whereas $-1$ is the worst punishment. Inside the layers targeted by the VTA, modulation of the output activity of each neural unit is achieved by simply multiplying the weighted sum of its inputs by the activation value sent by the VTA. In retrospect, this was not the best way to apply modulation to the different neurons. Indeed, in this context a neutral stimulus (one associated with an output activity for the VTA layer of 0) would completely extinguish all the neurons targeted by the VTA instead of leaving their activities unchanged as would be expected. A more interesting approach would have been to employ a scheme similar to the one discussed by Litt et al. (2006) in which rewards and punishments are either added to or subtracted from the activity of neurons, respectively (this type of scheme has also been considered by others beside, such as Frank (2005) and Gurney et al. (2001a, 2001b)).

### B.1.2 Re-implementing the ProtoEmo architecture

When the **ProtoEmo** architecture was included into **PrimEmo**'s design small changes in the connection pattern between the layers making up **ProtoEmo** were implemented. The projection from the lateral nucleus of the amygdala (LA) to the intercalated neurons (Itc) has been removed, while a bidirectional connection between the hypothalamus and the basal nucleus of the amygdala (BA) has been added. These changes however small they may seem are not without consequences.

The Itc has been suggested to implement a habituation mechanism providing feedfoward inhibition to the central nucleus of the amygdala (CA; see Section 5.1, Sub-Section A.1.4, as well as Whalen and Phelps (2009)). By having the hypothalamus send its activity to the BA, it shifts the responsibility of integrating sensory data describing the external environment with information about the individual from the CA to the BA. Furthermore, now that only the BA projects to the Itc, it means that the exact same information reaches both the Itc and CA. Consequently, the Itc can more precisely predict by what amount to inhibit the CA to silence its neurons.

In Section 6.1 it was suggested that the output activity of the BA could be interpreted as a novelty signal. Therefore, an unpredicted stimulus triggering a spike in BA's activity will be transmitted to both the Itc and CA. At first, and due to the Itc having a slower learning rate than the CA, neurons from the CA are potentiated, essentially raising the alarm for the rest of the brain. However, if the same stimulus is repeated for long enough, the Itc will slowly adapt its inhibitory activity to silence the CA preventing it from interfering with ongoing cognitive processes. In this context, the projection from the BA targeting the hypothalamus can be construed as the BA warning the control center of the autonomic nervous system about the presence of something potentially harmful. It is then the responsibility of the hypothalamus to prepare the individual for a fight-or-flight response.

### B.1.3 Connecting the models together

Since **PrimEmo** is the architecture that results from the aggregation of **ProtoEmo**, the PVLV and the basal ganglia, some decisions had to be made when connecting those models together. Although Section 6.1 already describes the overlapping areas and their meaning within **PrimEmo**, this section goes over the connections between the models in more details again.

For the PVLV, in both its original (O'Reilly et al., 2007) and updated (Hazy et al., 2010) forms (see Sub-Sub-Section 3.3.2 for a detailed description), the **ProtoEmo** architecture can be construed as an extension. Indeed, while in the PVLV only the central nucleus of the amygdala is included as the excitatory component of the '*Learned Value*' (LV) system, **ProtoEmo** is modeled after a survival circuit, comprised of the amygdala, thalamus, and the hypothalamus, found in the brain of rats and non-human primates. Consequently, extending the PVLV with **ProtoEmo** increases the explanatory power of the PVLV. To connect the two models together the following projections are required:

- From the lateral hypothalamic area (LHA) and the central nucleus of the amygdala (CA) to the pedunculopontine tegmental nucleus (PPT): the function fulfilled by those two indirect connections from both the LHA and CA to the VTA via the PPT is not clearly understood yet (Hazy et al., 2010; O'Reilly et al., 2007). According to both O'Reilly et al. (2007) and Hazy et al. (2010), activation of the PPT nucleus results in the dopaminergic system firing bursts of activities, but only for dopamine neurons that are already tonically activated. Additionally, because the PPT nucleus sends both cholinergic (inhibitory) and glutamatergic (excitatory) efferents to the dopamine system, it could also be involved in suppressing the phasic activity of the dopaminergic system when the triggering stimulus is sustained (Hazy et al., 2010).

- Directly from the CA to the ventral tegmental area (VTA): In the **PrimEmo** architecture the VTA stands in for the aggregation of the dopaminergic (DA) and serotonergic (5-HT) systems. The DA system is usually considered as being made of the VTA and the substantia nigra pars compacta (SNc). The 5-HT system mostly consist in the dorsal raphe nucleus. Within the PVLV model, the CA is hypothesized to be the excitatory component of the LV system. As such, the output activity of the CA is interpreted as a value attached to any stimulus that is reliably paired with a reward (Hazy et al., 2010; O'Reilly et al., 2007). In the **ProtoEmo** architecture, the CA has been hypothesized (see Chapter 5) to detect stimuli relevant to the individual's survival. Thus, the CA's output can be construed as an alarm signal, and the responsibility of determining the rewarding or punishing nature of a stimulus falls with the VTA. Even though it might seem otherwise, considering that what is relevant to an animal's survival is usually perceived as being either rewarding or punishing as well, the points of view expressed so far are compatible with each other (Rolls, 2007, 2013). Furthermore, Hazy et al. (2010, p.709) remarks that the CA is also active for innately rewarding stimuli (US), and learns to fire for an associated CS via conditioning (see Chapter 4 and Armony et al. (1995) for a computational model describing this mechanism). This observation lends support to both hypotheses at the same time.

- Between the basal nucleus of the amygdala (BA) and the ventral striatum (VS): In the **PrimEmo** architecture via its connections to both parts of the ventral striatum (Whalen & Phelps, 2009), as well as the other two components of the PVLV system, the BA sends a consolidated view of both the environment and the individual. Consequently, this allows the '*critic*' portion of the **PrimEmo** architecture (see Section 3.3 for a detailed description of the Actor-Critic as a mathematical model and its mappings to the animal's brain) to learn and predict the rewarding value associated with a global state. This is further supported by the fact that, even though, both BA and CA receive projections from different cortical areas, afferents to the BA are more numerous and of a finer granularity (Aggleton, 1992; Hazy et al., 2010; Whalen & Phelps, 2009).

- And connections from the VTA modulating the activities of the LHA and the CA: Assuming that the aggregation of the PVLV model with the **ProtoEmo** architecture is the '*critic*' portion of **PrimEmo**, then the output of the VTA would be equivalent to the TD error defined in Sub-Section 3.3.1. This discrepancy between predicted and actual reward is used by both the critic itself and the actor to adjust the strength of their synapses. Consequently, the output activity of the VTA is a learning signal sent to all components of the PVLV architecture.

In **PrimEmo**, consistent with the Actor-Critic model described in Section 3.3 the interactions between the '*critic*' and '*actor*' portions of the architecture are kept to a minimum. Only the two following projections connect the PVLV model with the basal ganglia:

- The DA system modulates both sub-populations of the dorsal striatum (DS): As described in Sub-Sub-Section 3.3.2 this modulatory connection between the VTA and the dorsal striatum is necessary to ensure that the basal ganglia (BG) can perform its role as an action selection mechanism. Moreover, in the context of the mathematical description of the Actor-Critic model in Sub-Section 3.3.1, the signal between the VTA and DS can also be interpreted as the discrepancy between the predicted and the actual reward received (the TD error) in a given state after performing the previously selected action.

- The motor cortex (MC) targets both sub-populations of the ventral striatum (VS): This second connection sends back the action the individual has just performed to the ventral striatum, and as will be detailed later to the BA as well. Since both targets of this projection are within the critic portion of the **PrimEmo** architecture it means that the TD errors are computed using $(state, action)$ pairs.

Similar to the PVLV, the basal ganglia and the **ProtoEmo** architecture do not overlap and very few connections link them together. However, the projections that do exist between those two structures greatly enhance the capabilities of **ProtoEmo**. The links implemented in the **PrimEmo** controller and their suggested function are as follow:

- The basal nucleus of the amygdala (BA) is target by both the pre-frontal (PFC) and motor (MC) cortices: In the PBWM model introduced by Frank et al. (2001), the PFC is shown to be the ideal candidate for fulfilling the role of working memory. Indeed, assuming that information is stored within the activation pattern of the PFC's neurons, the many recurrent loops that make up the PFC allow it to actively sustain any firing pattern over delays. Rapid update of the data stored inside working memory relies on the gating mechanism implemented by the basal ganglia. Whenever the gate is open, information is allowed to flow into the PFC and, therefore, modify the content of working memory. On the contrary, when the gate is closed new inputs are prevented from modifying the content of working memory, which is actively maintained by the PFC. Consequently, by connecting the BA with both the MC and PFC, the BA has access to the action being carried out, as well as high-level information concerning the current situation. Thus, the BA can integrate any incoming stimulus into a broader context, and, therefore, build a broader picture encompassing the state of both the environment and the individual, along with predictions about the future.

- The central nucleus of the amygdala (CA) receives a projection from the somato-sensory cortex (SSC): This connection is part of the model of the dual pathways to the amygdala suggested by LeDoux (1986) and Romanski and LeDoux (1992). As described at length in Chapter 4 and by Armony et al. (1995), this indirect pathway from the thalamus to the amygdala provides the CA with fine-grained data related to incoming stimuli, hence, allowing animals to discern between two adjacent frequencies for example.

Although, in animals the motor cortex (MC) is the brain area responsible for executing any planned action, in the **PrimEmo** architecture an output layer has been added on top of the MC. As introduced in Sub-Section B.1.1 areas belonging to the actor portion of **PrimEmo** are implemented as groups of layers, where each layer represents the program and salience associated with one of the actions available to the hero (see Sub-Sub-Section 3.3.2, as well as Redgrave et al. (1999) for a description of the '*dual population coding*' model). The motor cortex is no exception to this rule. Each unit in the output layer is fully-connected with a single layer from the motor cortex. The activity of a unit is computed as the average activity of its input layer. A softmax activation function (defined in Equation A.4) is then applied to the whole output layer to implemented a graded version of the winner-takes-all mechanism. Then, to decide which action to perform next the **PrimEmo** controller uses a greedy policy, which dictates that the action associated with the most activated unit is always selected. It should be noted that, due to the softmax activation function the values of the output units could have been interpreted as probabilities for their respective action to be selected as is the case for the asynchronous actor-critic controller described in Section 3.3.

Last but not least, for any of the recurrent connections present in the **PrimEmo** controller, such as between PFC and BA, or between CA and LHA, an exponential decay mechanism has been implemented. As a result, only 90% of an input unit's activity is transmitted to its targeted area via a recurrent projection. In animals, such a decay mechanism has been found in the PFC which has been suggested to be maintaining and updating the content of working memory (Frank et al. (2001); Shanahan (2010); Sloman (1999) to name but a few examples). Moreover, since the **PrimEmo** controller lets the network settle before observing the output, having a decay mechanism for any recurrent connection prevents any pattern of activity from being maintained forever causing the network's output to never stabilize.

In an effort to keep this appendix concise yet accurate, and due to the number of brain areas involved, the sizes of the layers making up the **PrimEmo** controller (shown in Figure B.1) have been defined in Table B.1.

Table B.1 *This table details the size of the different layers defined for the implementation of the* ***PrimEmo*** *controller. It should be noted that, wherever brain areas have been implemented as groups of layers, each group contains* 5 *layers (one layer for each action available to the hero). As a result,* ***PrimEmo****'s neural network shown in Figure B.1 includes* 60 *layers in total.*

| Brain Area | Layer \ Group | Size |
|---|---|---|
| External Input | Layer 1 | 22 |
| Internal Input | Layer 1 | 2 |
| Reward Input | Layer 1 | 1 |
| LHA | Layer 1 | 8 |
| Hypot | Layer 1 | 10 |
| Thal_1 | Layer 1 | 40 |
| LA | Layer 1 | 20 |
|  | Layer 2 | 10 |
| BA | Layer 1 | 50 |
|  | Layer 2 | 25 |
| Itc | Layer 1 | 12 |
|  | Layer 2 | 7 |
| SSC | Layer 1 | 20 |
| CA | Layer 1 | 30 |
|  | Layer 2 | 15 |
| PPT | Layer 1 | 10 |
| VS_1 | Layer 1 | 20 |
| VS_2 | Layer 1 | 20 |
| VTA | Layer 1 | 30 |
|  | Layer 2 | 1 |
| DS_1 | Group 1 | 20 |
| DS_2 | Group 1 | 20 |
| GPe | Group 1 | 15 |
| STN | Group 1 | 15 |
| SNr | Group 1 | 20 |
| Thal_2 | Group 1 | 20 |
| PFC | Group 1 | 20 |
| MC | Group 1 | 20 |
| Output | Layer 1 | 5 |

## B.2 Evolving PrimEmo for optimal survival and decision-making

The training process for the **PrimEmo** controller is very similar to the one described in Section A.2 for the **ProtoEmo** controller. A simple genetic algorithm (GA) is employed to update the synaptic weights and optimize **PrimEmo**'s strategy for gathering as much gold as possible while surviving in Viking Doom (a game introduced in Section 6.3 and detailed at length in Section C.2). The reasons for employing a GA once again are similar to those expressed at the beginning of Section A.2. Not only does using a GA draw nice parallel between animals evolving emotions and now virtual agents evolving emotions of their own, but it is also highly parallelizable, and, therefore, more computationally efficient (even more so when compared to alternatives such as backpropagation through time (Lillicrap et al., 2016; Werbos, 1990) which would require extensive amounts of memory and compute power). Due to its ease of use and the possibility it affords for splitting computation across different compute node, the Distributed Evolutionary Algorithms in Python (DEAP; written by Fortin et al. (2012)) library was also used again for implementing the training process.

The remainder of this section is organized in three parts. The first Sub-Section introduces the population and individuals used for the genetic algorithm. The next part describes step by step the protocol followed for evaluating and evolving the individuals as part of the training process. Finally, the last Sub-Section presents the evolutionary operators implemented for selecting the parents, as well as mating them to produce the next generation.

### B.2.1 A population of synaptic strengths and biases

As with any other artificial neural network the means by which the **PrimEmo** controller learns to perform a task is through the update of the strengths of the synapses connecting units together. Consequently, the individuals defined for the GA are based on the weight matrices associated with the different connections between layers. On the contrary to the training process implemented for the **ProtoEmo** controller, only a single population was used instead of multiple co-evolving ones (see Sub-Section A.2.1 a definition of the concept of cooperative coevolution in the context of the DEAP library). Consequently, each individual is a list containing the matrices for all the projections defined in **PrimEmo**. The biases for all neural units are initialized to $b = 0.1$ and left unchanged for the whole duration of the experiment. Since the implementation of the **PrimEmo** network relies on the Keras framework (Chollet,

2015) backed by the Tensorflow library (Abadi et al., 2015), then a weight matrix is represented as:

$$\mathbf{W} = \begin{pmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,j} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i,0} & w_{i,1} & \cdots & w_{i,j} \end{pmatrix} \tag{B.3}$$

The width of the matrix $j$ is set to the size of the layer and the height $i$ to the number of units in the sending layer(s). As a result, an individual takes the following shape:

$$ind = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n] \tag{B.4}$$

Where $n = 57$ corresponds to the number of layers implemented for the **PrimEmo** controller. Finally, at the beginning of the training phase the matrices inside each individual are initialized with random values sampled from a Gaussian distribution with a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$. The population used for training the **PrimEmo** controller to play Viking Doom contained 50 individuals.

### B.2.2 Evaluating individuals

By definition genetic algorithms search the space of available solutions for the optimal set of parameters for a given problem, be it a task or a mathematical function. To aid in this endeavor each individual is first evaluated and assigned a value representing how fit it is for a specific problem. This '*fitness*' value is then utilized by the different evolutionary operators to select parents and mate them to build the next generation. Although, it is possible in the DEAP framework to define multiple fitness corresponding to different objectives, and assign weights to each fitness to denote the relative importance of each objective, the **PrimEmo** controller is only alloted a single goal, that is to collect as much gold as possible before the end of a game. Furthermore, due to the randomness inherent in the design of Viking Doom (see Section 6.3 and Section C.2 for more details) the fitness of each individual is set to the total amount of gold gathered on average over 10 games. The fact that only one objective related to the average total amount of gold was defined has been done for two reasons. Firstly, it provides a fair context in which to compare **PrimEmo** with the other two reinforcement learning based algorithm. More importantly, however, is that the intention behind this experiment was to have emotions naturally emerge from the evolutionary process rather than be the consequence of some external constraints.

For the genetic algorithm implemented to train the **PrimEmo** controller, the evaluation phase can be split into the following steps:

1. An empty template embodying the design of the **PrimEmo** architecture (described in Section 6.1 and further detailed in Section B.1) is pre-loaded.
2. The matrices contained within the individual are utilized to set the synaptic strengths and biases of all connections between layers.
3. The completed neural structure is then evaluated by having the **PrimEmo** controller play 10 games of Viking Doom. Each game of Viking Doom lasts at most 1000 cycles or until the hero dies.
4. At the end of each game, the total amount of gold gathered by the hero is recorded. When all 10 games are completed the average amount of gold gathered is then assigned as the individual's fitness.
5. Finally, the next individual is retrieved and the process repeated from step 1 until the whole population has been evaluated. From within the evaluated population the parents are selected and mated to produce the next generation.

The results analyzed in details in Section 6.4 were obtained by training the **PrimEmo** controller for 89 generations only. Given that the population size was set to 50 individuals, it means that a meager 4450 sets of weight matrices were evaluated.

### B.2.3 Making the next generation

After the evaluation phase is complete, the second step of a standard genetic algorithm is to evolve a select number of parents to produce the next generation. This step can be split into three main operations. Starting with selecting suitable parents. Although, selection mechanisms usually determines an individual's probability of being chosen based on its fitness, and ultimately decide whether to pick it as a parent or not, in the GA implemented to train the **PrimEmo** controller the fitness does not intervene and is only used to rank individuals within the population. The '*rank selection*' algorithm, which is the standard name for the selection mechanism employed in this experiment, applies constant pressure by favoring higher ranked individuals. This is a desired characteristic especially at the beginning of the learning process when all individuals have similar fitnesses. It ensures that the next generation stems from the best parents, while still maintaining a diverse genetic pool.

Once appropriate parents have been selected, the offspring population is created via a crossover operation which combines parts from two parents to produce two children. Given that individuals have been defined as lists of weight matrices, the GA implemented for the training process applies a modified version of the standard '*two-points*' crossover. In this modified version, and because each matrix represents a different synaptic connection, the standard two-point crossover is applied row-wise between corresponding matrices from both parents. The probability for two individuals to be combined is set to 0.8.

The third, and final, evolutionary operation applied to the resulting population of children is mutation. Since individuals are lists of real valued weight matrices, mutation within the implemented GA takes the form of random replacement with values sampled from a Gaussian distribution of mean $\mu = 0$ and standard deviation $\sigma = 1$. The probability for an individual to be mutated is set to 0.05. Furthermore, if an individual is selected for mutation, then the probability for an element of a weight matrix to be replace is set to 0.8.

In the end, after evaluating the whole population, then ranking and selecting the parents for the next generation, the selected individuals are evolved through combination and mutation to produce offspring. The old population is then replaced with the children, and the genetic process starts all over again for many generations.

# APPENDIX C

# Specifics of the Viking Doom project and the second version of the game engine

Viking Doom is a benchmark in the shape of a roguelike game. It is used in Chapter 6 to investigate the role of emotions in the decision-making process through the *PrimEmo* architecture described in Section 6.1 and Appendix B. Furthermore, the use of a game as a benchmark is all the more convenient, since reinforcement learning agents (more specifically a deep Q-learning (DQN) and an Advantage Actor-Critic (A2C) controllers) have been implemented for comparison. Indeed, games by their very nature have a dense reward function (usually this is the current score), which can be used to efficiently direct the learning process. Therefore, even if the nature of the task is quite complex both the DQN and the A2C agents should be able to relatively quickly learn an optimal strategy, while avoiding divergence or local optima.

As explained at length in Section 6.2 and Section 6.3, Viking Doom was first intended to be used in a competition. As such, an accompanying website and a JavaScript backed display have been implemented. However, because the competition did not yield any useful results and the environment's design was flawed a second, standalone, version of Viking Doom was created. This appendix describes in more details the interface developed for building agents capable of playing the version of Viking Doom. It then provides all the specifics required for implementing the second version of Viking Doom's game engine.

## C.1 Defining a protocol for players to communicate with Viking Doom's game engine

When organizing an on-line machine learning competition it seems there are two popular approaches. The first one is to provide the game engine in advance, so that players can develop their virtual agents. For the competition, participants then upload the final version of their agents to a central repository. It is then the responsibility of the server hosting the competition to automatically execute the different agents, and have them compete in a round-robin tournament. This kind of scheme puts all the cost and complexities on the competition's host. Furthermore, it also limits the programming languages in which players can implement their virtual agents.

Another possibility, and the one adopted for the first version of Viking Doom, is to design the game around a client/server model. This not only spreads the cost between the host and the participants, but also alleviates some of the problems faced by the previous approach. Firstly, it does not force participants to use a single programming language to implement their virtual agents. Since HTTP is a standard protocol, clients can be written in virtually any language. Furthermore, since participants do not have to submit their virtual agents to be run on a central host computer, it removes any need for automation scripts, and allows players to more quickly deploy new versions of their programs. However, the flexibility afforded by a client/server architecture comes at a cost. The most important, in the point of view of the participants, is that it requires players to have a '*good*' (in terms of speed and stability) Internet connection and to leave their computers running for the whole duration of the competition. Furthermore, since different programming languages can be used to implement virtual agents, it means that the server has to send data in a format understandable by all. This section, therefore, details the design of the game engine and explains how it mitigates some of the problems expressed so far to provide an even playing field for all participants.

### C.1.1 Registering a player

Before any adventuring business can begin, users first have to register as new players. To do that they need to provide a '*username*' and some personal information. Once the subscription form has been submitted to the server each user receives a unique code made of 15 alphanumeric characters. This code is the only way the server can identify heroes and link them to a specific username. Consequently, if a user ever loses their code they will have to register a new player and restart their journey through the ranks.

Now that a new players has been created and associated with unique codes, users can finally start programming their agents. Indeed, Viking Doom is not a game to be enjoyed by mere humans. It pitches AIs (artificially intelligent agents) against AIs and display the resulting battle on the associated website's home page for humans to enjoy (see Figure C.1 for some examples of this display; a live version is also available as a video [1]). To help in the process of creating an agent, users can have a look at the random agent available in the VikingDoom project on Github [2]. There they should also find the game engine and its accompanying website. This will allow them to test their agents before entering the real battle. By cloning or forking the whole project, users are free to further improve the game and use it for their own end. As shown in Figure C.2, an agent at its core acts like a simple HTTP Client sending '*PUT*' and '*POST*' requests to the Viking Doom server and receiving a JSON formatted response. It is then up to the user to include some logic in the agent and have it devise some brilliant strategies to always gather more gold, while avoiding danger. Even though, the game engine (acting as the server) and the website are written in Python, agents can be implemented in any programming language, be it JAVA, C or C++. JSON has been chosen, because it is a language-independent human-readable data format commonly used to store and transmit objects consisting of attribute-value pairs. Even though, it is derived from JavaScript, many programming languages are able to generate and parse JSON formatted data. Furthermore, although textual data is not the most efficient way (in terms of bandwidth requirements) to transmit information over the Internet, JSON formatted information is easy to handle for clients and does not incur as much overhead as data compression/decompression would. Finally, the different objects (detailed in Sub-Sub-Sections C.1.3 — C.1.3) have been designed to contain the strict minimum required to describe the state of the environment and of all the heroes. Further optimizing bandwidth usage and communication speed.

### C.1.2 Joining a game of Viking Doom

To begin a new game, well players simply ask for it by sending a '*PUT*' request to the url: `http://www.vikingdoom.com/game/new/your_code_here/`, where '*your_code_here*' has to be replaced by the code given to each user upon the creation of their associated player. As an answer the server sends a JSON formatted game object (an example of which is shown in Listing 1) back. The game object (described in more details in Sub-Sub-Section C.1.3) contains

---

[1] `https://youtu.be/yndgfQ2HobM`

[2] `https://github.com/casparluc/VikingDoom`


Now Playing
2485
27
0
66
4230
98
6045
69
Now Playing
0
99
0
99
0
100
0
100

Figure C.1 Examples of the display used on the home page of the website associated with the first version of Viking Doom.

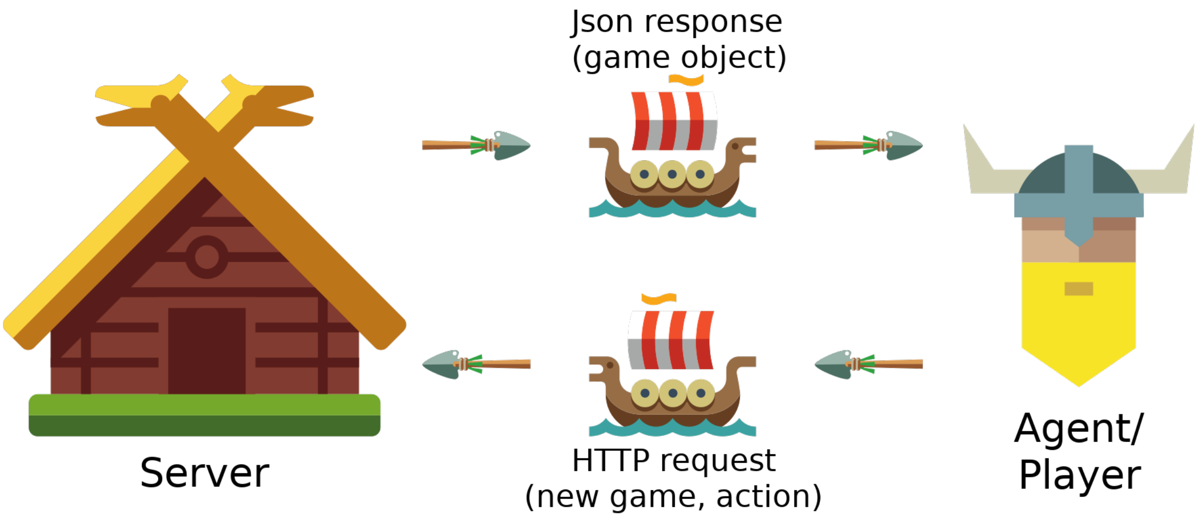


Figure C.2 A graphical representation of the communication protocol between the player (the client) and the game engine (the server).

all manner of information, but the most important one in this case is the '*URL*' attribute. This contains the url at which players send '*POST*' requests to move their hero around.
Before going further into playing the actual game, users have to be aware that upon sending the initial '*PUT*' request, the server can send back one of the three following HTTP status code:

- '*201_CREATED*': the corresponding player has been found and added to a game.
- '*409_CONFLICT*': a player associated with the current user is already part of this game. Since, users can only have one player per game, it means they will have to ask to be added to another game.
- '*404_NOT_FOUND*': the server was unable to find a player linked to the unique code the user sent.

### C.1.3 Playing a game of Viking Doom

Playing the actual game follows a pattern similar to that of asking for a new game. Users simply send the action, they want their heroes to perform next, through a '*POST*' request and wait for the server to send the JSON formatted game state back. The same two steps are repeated until all players are '*Terminated*' or the game is '*Finished*'.

When asking for a new game users receive and should parse the JSON formatted response. As mentioned earlier, the game object contains the '*url*' to which users should address all of their consequent '*POST*' requests for the up coming game. The body of the request should contain:

- code (required): the 15 alphanumeric characters corresponding the player.
- action (required): the action the hero should to perform next. See Sub-Sub-Section C.1.3 for the different values this parameter can take.

Shortly after the server sends back a game object, that users can parse to take decisions about their next move. Users only have one second to send the next action and it has to be a valid move. Failure to do so more than three times leads to the linked player being '*Terminated*'. A terminated player can still still send actions, but none will be considered by the game engine. If everything goes well and all agents manage to play for 1000 turns, the game enters its '*Finished*' state. Scores for the game are recorded in a database and combined with previous scores to form the global rank of each user. Once a game is '*Finished*' actions can no longer be sent to the server. Similar the previous phase, sending an action in a '*POST*' request can result in four different HTTP status codes:

- '*504_GATEWAY_TIMEOUT*': the game is still waiting for more players to join. Or the previous game is not finished yet and this one is still in the queue.
- '*503_SERVICE_UNAVAILABLE*': the game the user trying to access is '*Finished*'. No one is playing anymore and the scores have been recorded. Users should ask for another game.
- '*404_NOT_FOUND*': the server could not find the game or player associated with the code sent by the user. The user should make sure that he is sending the correct code in the body of the '*POST*' request and that the '*url*' is correctly formatted.
- '*200_OK*': the associated player has been found and the game is still playing. The user's request for the next action has been sent to the game engine and will be considered shortly.

### The game object

The game object is the top level container. It includes information about the map, all heroes, the hero attributed to the current player in particular and the aforementioned '*url*'. Here is an example of a game object:

The different fields are:

"Game object"

```
{
   "turn": 0,
   "url": "http://www.vikingdoom.com/game/play/XKQUD3W620ZJRB4/",
   "code": "XKQUD3W620ZJRB4",
   "state": "P",
   "heroes": [ ... ],
   "map": { ... },
   "your_hero": { ... }
}
```

Listing 1 *An example of a JSON formatted game object returned by the Viking Doom server.*

- ‘*map*’: a map object, containing information about the state of the environment. See Listing 2 for more details on its content.
- ‘*players*’: in each game four players are participating. This list of player objects contains the heroes’ state (health, gold, position and more). Refer to Listing 3, as well as Sub-Sub-Section C.1.3 below for an in-depth description of its content.
- ‘*turn*’: this is a simple integer indicating the game turn to which this object corresponds to. The turn meter begins at 0 and goes to a maximum of 1000.
- ‘*url*’: the ‘*url*’ to which users have to send all their actions in the shape of ‘*POST*’ requests.
- ‘*code*’: similar to the code attributed to each player, this string of 15 alphanumeric characters is an identifier for this game.
- ‘*state*’: a single character representing the game’s state:
  - ‘*I*’: Initializing
  - ‘*W*’: Waiting
  - ‘*P*’: Playing
  - ‘*F*’: Finished
  - ‘*T*’: Terminated
- ‘*your_hero*’: A hero object that repeats the state corresponding to the hero the current player owns. It is a shortcut for accessing what really matters to the agent.

### The Map object

The map object includes information relevant to the environment that players are currently walking, fighting and dying in. Here is a close up of the map object alone: The fields present in this object are:

"Map object"

```
{
   "str_map":
   "##          !S?B              ~U


    $_      ~P            !S          ##
                  @2                  ##
                  ######              ##
      ##        ##########      $_  @3##
      ##        ##########            ##
    ##          ##########            ##
    ##      ?G    ########      ?P    ##

                                      $_
    $_          $_            !S
              ########
                                  ##
        !S                @1 @4

           ?P    !D          $_$_  $_   ",
   "width": 18,
   "height": 18
}
```

Listing 2 *An example of a JSON formatted map object returned by the Viking Doom server.*

- str_map: a string representation of the current state of the map.
- width: an integer representing the width of the map in columns. Equivalent to the size of the map along the X axis.
- height: an integer representing the height of the map in rows. Equivalent to the size of the map along the Y axis.

The map object contains an encoded representation of the current environment's state. For each turn agents have to decode this string and extract any relevant information. Each object is represented by two characters as follow:

- ‘*##*’: a simple wall. Heroes cannot step on such tiles,
- ‘*$_*’: a mine that does not belong to anyone,
- ‘*$1*’: a mine that belongs to the first player,
- ‘*@1*’: the hero controlled by the first player,
- ‘*!S*’: an enemy unit:
    - ‘*!S*’: references skeleton units,
    - ‘*!D*’: represents The Dragon,
- ‘*?P*’: an item:
    - ‘*?P*’: for potion,
    - ‘*?G*’: for a purse full of gold,
    - ‘*?B*’: for a chest of gold,
- ‘ *P*’: indicate the location and type of a market. With:
    - ‘ *P*’: for potion selling markets,
    - ‘ *U*’: for strength upgrade caravans.

It should be noted that Orc units are not represented in the encoded map since each mine has an orc guarding it. Therefore, players can assume that: “*Where there is a mine, there is an orc*”.

### The Hero object

Inside the global game container, there are five hero objects. One for each player and a shortcut to the hero associated with the player originating the request (aptly named ‘*your_hero*’). These objects describe the state of each hero as seen in the following example: Where:

- ‘*hero_id*’: an integer in the range $[1,4]$ uniquely identifying the hero in the current game,
- ‘*pos_x*’: is the current position of the hero along the X axis,

"Hero object"

```
{
   "hero_id": 1,
   "pos_x": 11,
   "pos_y": 15,
   "spawn_x": 7,
   "spawn_y": 14,
   "health": 100,
   "gold": 0,
   "strength": 1,
   "state": "P",
   "action": "E"
}
```

Listing 3 *An example of a JSON formatted hero object returned by the Viking Doom server.*

- '*pos_y*': is the current position of the hero along the Y axis,
- '*spawn_x*': is the initial position of the hero along the X axis,
- '*spawn_y*': is the initial position of the hero along the Y axis,
- '*health*': an integer representing the hero's level of health,
- '*gold*': the amount of gold amassed since the beginning of the game,
- '*strength*': indicates the level of strength the hero currently has,
- '*state*': a single character representing the player's state:
  - '*I*': Initializing
  - '*P*': Playing
  - '*T*': Terminated
- '*action*': a single character representing the action performed by the hero during the last turn. In addition to the available actions described below, this fields can take the values:
  - '*F*': Fight
  - '*B*': Buy

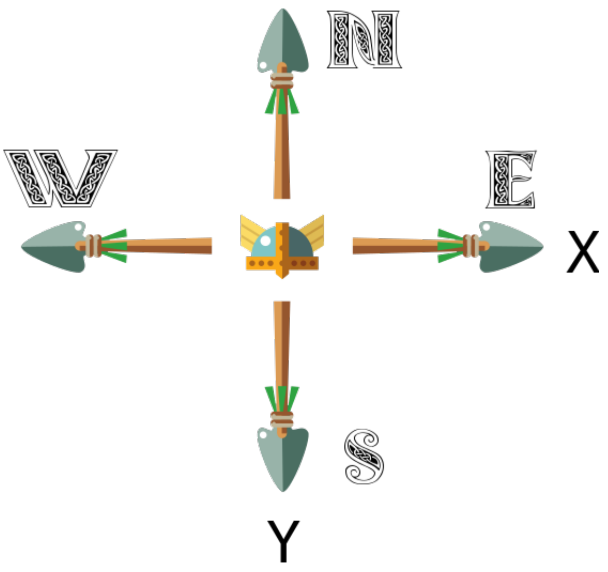


Figure C.3 A graphical representation of the different actions a player can request its hero to perform.

Even though, they are called actions the request players send for their hero is more of a direction for the hero's next move. Thus, as the picture above shows in the '*POST*' request the '*action*' variable can contain only one of the following characters:

- '*N*': going upward along the Y axis,
- '*W*': going left along the X axis,
- '*S*': going down the Y axis,
- '*E*': going right along the X axis,
- and finally '*I*': stands for idle, if the player wishes for his hero to remain standing on the same tile.

## C.2 Viking Doom 2.0: specifics regarding the game engine

At the end of Section 6.2 a peculiar strategy adopted by all the agents was brought up. This strategy is characterized by the hero either not moving at all for the whole game, or moving in one direction until the hero bumps into a wall and then stays in place. Throughout Section 6.3 possible causes for learning algorithms to fall into this local optimum were briefly discussed:

- Since heroes could not step on an already occupied tile, they could only gather items or fight enemies on neighboring tiles. A hero's neighborhood was defined as the set of tiles above, below, left and right of its current location. Consequently, on a single game cycle a hero could gather up to four items and/or fight up to four enemies. Combined with the fact that items were randomly scattered around the map, it means that heroes had a high probability of picking up gold while staying idle.
- Once a hero had gained ownership of a mine, its level of gold would rise by a constant amount on each game cycle. Since the reward's value heavily depends on the hero's level of gold, it means that any action taken would be positively reinforced, hence, confusing the learning algorithm.

Additionally, the first version of the game engine was meant to work as a Web server accepting and answering requests from remote players on the Internet. This client/server architecture was dictated by the on-line nature of the competition. However, moving away from the on-line competition and toward a standalone experiment meant that the game engine now had to run on a High Performance Computer. Consequently, the Web server had to be abandoned. To solve all those issues it was,therefore, necessary to redesign the environment from the ground up and establish new rules for the game. Enters Viking Doom 2.0.

This section is a complement to Section 6.3, and provides all the details required for implementing Viking Doom. Alternatively, it is also possible to clone or fork the Viking Doom project on GitLab [3], and run the experiment presented in this thesis on your computer.

### C.2.1 Configuring a complex and dynamic environment

As already mentioned in Sub-Section 6.2.2, the background and obstacles of every maps in the first version of Viking Doom were hand drawn. This was a tedious process, which limited the number and variety of environments available. To streamline the map creation process, in Viking Doom 2.0, configuration files have been introduced. As is shown in Listing 4, this file not only contains the width and height of the map, it also defines the position of all the obstacles, as well as areas and their associated quotas. Finally, at the end are specified the maximum number of items and enemies, on a per type basis, that can appear at once in the environment. Using configuration files, rather than static images, not only allows for quickly designing new and interesting environments, a great asset for the experimentation and exploration process, it

---

[3] `https://gitlab.com/davinellulinvega/vikingdoomv2.git`

means that the creation process could also be entirely automated. Regardless of how the files are produced, for running an experiment it is required to provide a list of configuration. To train the different controllers, it is then possible to either go through the list of environment sequentially or randomly. Both are valid choices, but have different advantages and drawbacks. Considering the environments randomly is more suitable for training a controller, but means that reproducing the results might be more difficult. The reverse being true for going through the different configurations in sequence.

At the beginning of each game, a configuration file is read and the environment is initialized with the corresponding set of obstacles, enemies and items. As mentioned in Sub-Section 6.3.2, those objects and the heroes are contained and managed by the environment. This means that whenever an item or enemy disappears from the map, it is the environment's responsibility to replace it on the next turn. Similarly, the attributes from all enemies and heroes currently on the map, are also adjusted by the environment.

"Environment configuration"

```
{"width": 10,
"height": 10,
"obstacles":  [[2, 2], [2, 7], [7, 2], [7, 6]],
"areas": [
   {"id": "0", "position_min": [0, 0], "position_max": [4, 4]},
   {"id": "1", "position_min": [4, 4], "position_max": [9, 9]},
   {"id": "2", "position_min": [0, 4], "position_max": [4, 9]},
   {"id": "3", "position_min": [4, 0], "position_max": [9, 4]}
],
"quotas": {
  "0": {"hero": 1, "enemy": 25, "item": 18},
  "1": {"hero": 1, "enemy": 25, "item": 18},
  "2": {"hero": 1, "enemy": 25, "item": 18},
  "3": {"hero": 1, "enemy": 25, "item": 18}
},
"max_enemy_counts": {"skeleton":  0, "troll":  0, "dragon": 2},
"max_item_counts": {"purse":  0, "chest":  12, "potion":  12}}
```

Listing 4 *An example of a JSON formatted configuration file, from which a complete environment can be created.*

### C.2.2 Assigning values to the incentives and dangers

In Sub-Section 6.3.3 six objects are described as populating the environment, aside from obstacles and heroes. Those objects can be sorted into two main categories: collectible '*items*' and '*enemies*' to fight. In this version of Viking Doom heroes can collect the following items:

- *Purse* and *chest*: purses only grant 5 units of gold, while chests are worth 25 units of gold. This discrepancy in reward has been included in the design, so that controllers could adapt their strategy depending on the expected reward. For example, if a hero finds itself in the following scenario: the closest gold item is a chest, but it is right next to a dragon, while the second closest item is a purse. It is expected of any controller, whose completed its learning phase, to chose to collect the isolated purse. The logic behind this choice is quite simple. On the one hand, going toward the chest (the highest reward) comes with the risk of having to fight an enemy (a dragon, the strongest unit) and maybe dying in the process. Picking the purse as the next target, on the other hand, is certain to yield a reward, albeit a smaller one.
- *Health potion*: Whereas purse and chest can be considered virtual incentives, since they are only relevant in the context of this game, health potion is a real motivator necessary to the hero's survival. Indeed, the only option for a hero that has taken some damage and is not at full health anymore, is to collect potions which increase its health level by 25 units each. It is expected for a controller, whose learning phase has been completed, to show a preference for health potions rather than chests or purses, whenever a hero is in need of healing.

To allow the investigation of both positive and negative emotions, Viking Doom also provides dangers in the form of enemy units. During a typical game heroes can come across three types of opponents. From the weakest to the strongest they are:

- Skeletons: Every instances of this type are initialized with 10 units of health and a single unit of strength. Given the low-level of health skeletons are granted, heroes are expected to discover that they can easily defeat them and therefore increase their strength. Indeed, killing a skeleton is rewarded with one unit of gold and one of strength. Similar to the first version of Viking Doom, skeletons can move on every game turn. Compared to the previous implementation, though, their movements are not random. Instead, upon initialization a skeleton is assigned an initial position and has an area of effect. This area is a circle around the unit's initial position and with a radius of one tile. Once initialized a skeleton unit, therefore, executes a very simple strategy. By default a skeleton waits for a hero to enter its area of effect. The enemy unit then uses an A* algorithm to determine the quickest path to the hero, and move in that direction. If the skeleton manages to catch up with the hero, they automatically fight. A skeleton will pursue a hero until the hero either exits the area of effect or dies. At this point the skeleton simply moves back to its initial position, and waits for another hero to come by. If two heroes are within the area at the same time, only the closest hero from the skeleton's current location becomes the target. Even though, skeletons are weak and should not be a threat to any healthy hero, a swarm can quickly become quite deadly.

- Trolls: They are characterized by a level of strength of 12.5 units, 25 units of health and the fact that they are static. A fully trained controller will rarely, if ever, meet a troll, since it should have learned to avoid enemies through its encounters with the skeleton units. Unless a hero is seeking a fight with a specific troll, it is expected that the hero will rather learn to optimize its way around the static troll. Trolls are only virtual obstacles, since they can be fought and defeated. Doing so grants the victor 10 units of gold and three units of strength.
- Dragons: With 75 units of health and 25 of strength, this enemy is the most fearsome. Similar to skeletons, dragons are mobile and are assigned an initial location. From this location a circular area of effect with a radius of two tiles is defined. Any hero setting foot in this area will fall prey to the dragon's wrath. Dragons' movements are also planned based on the A* algorithm and they retreat to their initial position when the hero dies or leaves the area. Unlike skeletons, dragons can only move on odd turns of the game. Their fighting ability is not impeded, only movement is.

No matter its type when an enemy unit is on the same tile as a hero, it automatically fights said hero dealing damage at every turn of the game. Furthermore, it should be noted that within the Viking Doom project, whenever distance is mentioned it is taken to be the Manhattan distance or L1 norm.

To address the issue of heroes being able to '*win*' the game without moving, any randomness in the initialization process of objects has been removed, and two rules common to all objects (heroes included) have been enforced. Firstly, heroes can only interact with objects on the same tile as them. That is, to collect items or fight enemies (or other heroes) a hero's position has to overlap with that of the object in question. This was introduced to decrease the probability of heroes collecting items without having to move. To actually force heroes to work before obtaining a reward, the second rule that has been implemented is: an object cannot be initialized on an already occupied tile. So, for example, items cannot appear on the same tile as heroes or enemy units.

To allow a more flexible design process of the map and a precise placement of the different items, enemies, and heroes, compliant with the rules above, the concepts of ‘*area*’ and ‘*quota*’ have been introduced. This is also a measure to prevent another phenomenon observed in the previous version of the project. Given that previous maps were bigger in size it was less apparent, but narrow areas (i.e.: bottlenecks) usually ended up being overcrowded, since not a lot of heroes were going through to collect items or battle enemies.

Areas and their associated quotas are a flexible solution to all those problems. As shown in Listing 4, both areas and quotas are defined in the configuration file. An area is simply a square shape zone characterized by a minimum and maximum position along the X and Y axis. It is not actually drawn on the map, but virtually constrains objects (such as heroes, items and enemies) to appear in certain parts. Due to the restriction on the shape, areas can overlap or be independent from each other.

Each area has three associated quotas: one regulating the number of heroes that can be initialized in this area, another for the number of enemies, whose initial position is within the area and a last one for the number of items present in the area at the same time. Note that the quotas are not specific to each sub-type of items or enemies. This is so that if an area already contains a dragon, for example, only a few trolls or skeletons can be added, but no dragon. As such it creates a system that is simple to configure, balances itself out and avoids having any overcrowded area on the map.

In addition to all the other attributes defined above, each item and enemy unit is also assigned a ‘*fulfillment rate*’. When adding an object to a specific area, the environment decreases the quota by the corresponding fulfillment rate. Deleting an object, therefore, increases the area’s quota again, allowing other items or enemies to be created. Consequently, whenever the environment creates a new item or enemy unit, it first ensures that the maximum number of this type of object has not been reached yet (this number is part of the environment’s configuration as seen in Listing 4). Then it goes through all the areas and checks which one has a quota higher that the object’s fulfillment rate. If more than one area meets this condition, an area is chosen at random, using the quotas of each area as weights for the sampling process. Finally, a random valid position is picked within the selected area. Here a valid position is defined as a tile, which is not currently occupied by an enemy, a hero, an obstacle or an item. The fulfillment rate of a hero, a potion, a purse and a skeleton is set to one. Trolls and chests both have fulfillment rates of five and dragons are the highest with a rate of 25.

### C.2.3 Implementing controllers for the heroes

This final part is an extension to Sub-Section 6.3.4 as it provides the specifics required for implementing the Deep Q-learning (DQN) and Advantage Actor-Critic (A2C) controllers. The source code for both controllers is also available on GitLab [4] as part of the Viking Doom project.

The main difference between Deep Q-learning (Mnih et al., 2013) and the Q-learning algorithm, detailed in Sub-Section 3.2.4, is that Deep Q-learning makes use of a neural network in place of an array to approximate the value function. Based on the features provided by the generic controller, the Deep Q-learning network outputs the Q-values corresponding to all actions at once. The vector of Q-values thus computed is then passed to the $\varepsilon$-greedy policy, which decides on the next action to perform. The specific implementation written for this experiment contains ten hidden layers, made of ReLU neural units, and a simple linear output. A linear output is used, since the neural activity of each unit represents the expected value of one of the available actions. The networks parameters (i.e.: the different weights and biases) are updated using the Adam optimizer (Kingma & Ba, 2015). At its core, this optimizer relies on the backpropagation algorithm to propagate the loss from the output layer to the input. For the output layer, the Mean Squared Error (MSE) is used as the loss function. In its standard form, the MSE is expressed as:

$$MSE = \frac{1}{n} \times \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \tag{C.1}$$

Where $n$ is the number of predictions (or the size of the output layer), $Y$ is the vector of observed values for the variable being predicted and $\hat{Y}$ the vector of predicted values. However, the ground truth for the different Q-values ($Y$ in Equation C.1) required by this standard definition of MSE is unknown. In fact, it the agent's goal to learn an approximation of these Q-values, while only being told how better or worse the new state is, compared to the previous one. What the reinforcement learning process provides though, via the TD error described in Equation 3.5, is an estimate of the true Q-values: $Y = r_{t+1} + \gamma \times v(S_{t+1}, a; \omega_t)$. It should be noted, that in Equation 3.5 the Q-value function was only dependent of the state, since in a classical

---

[4] https://gitlab.com/davinellulinvega/vikingdoomv2.git

conditioning experiment the subject is passive and receives a reward/punishment regardless of any action performed. For the Viking Doom task (which can be seen as an operant conditioning experiment), on the other hand, the agent is rewarded or punished for performing an action from a given state. Hence, the Q-value function depends on both the state $S_t$ and the action $a$. Putting it all together results in a loss function resembling the TD error of Equation 3.5, for the Deep Q-learning network:

$$loss_{DQN} = \begin{cases} \frac{1}{n} \times \sum_{i=1}^{n} (r_{t+1} - v(S_t, a_i; \omega_t))^2 & \text{if the episode ended} \\ \frac{1}{n} \times \sum_{i=1}^{n} (r_{t+1} + \gamma \times v(S_{t+1}, a_i; \omega_t) - v(S_t, a_i; \omega_t))^2 & \text{otherwise} \end{cases} \tag{C.2}$$

Furthermore, to avoid catastrophic forgetting and improve the neural network's stability, the Deep Q-learning controller makes use of both a target network, updated every 300 cycles, and experience replay.

The Advantage Actor-Critic (A2C) implemented follows the architecture suggested in Mnih et al. (2016). However, rather than having many asynchronous learners the current implementation only has a single agent. Hence, the lack of '*Asynchronous*' in the name. In its simplest version, the A2C architecture simply expands on the basic Actor-Critic model detailed in Section 3.3. The Critic, whose task it is to estimate the value of a given state, uses the same loss function as the Deep-Q learning network described above (see Equation C.2). The only difference being, that the Critic only learns to expect a reward irrespective of the last action performed. Therefore, the value function, the Critic seeks to optimize, only depends on the environment's state, and only a single linear neural unit is necessary to represent the Critic's output. Consequently, the equation corresponding to the Critic's loss can be rewritten as:

$$loss_{crit} = \begin{cases} (r_{t+1} - v(S_t; \omega_t))^2 & \text{if the episode ended} \\ (r_{t+1} + \gamma \times v(S_{t+1}; \omega_t) - v(S_t; \omega_t))^2 & \text{otherwise} \end{cases} \tag{C.3}$$

As determined in Section 3.3, the role of the Actor is not only to choose the next action to perform, following the softmax policy defined in Equation 3.10, but also to optimize said policy. To accomplish that the Actor updates its weights, at each time step, following the rules described by Equation 3.9.

This is where the A2C architecture really differentiate itself from the standard Actor-Critic model imagined by Sutton (1984). In the original version, the update rule for the Actor's weights did not include the TD error ($\delta_t$), instead relying on the Q-value associated with a given state to compute the change. Getting back to Equation 3.9b, the rule would be closer to:

$$\theta_{t+1} = \theta_t + \varepsilon \times v(S_t;\omega_t) \times \mathbf{z}_t^{\theta} \tag{C.4}$$

However, this expression presents a serious problem. For example, if the estimated values, for three states: A, B and C, are close to one another, such as 102, 103 and 101, respectively. Going from the current state to any of those three would result in a considerable update to the network's weights, but most of the information important to the learner is lost. Indeed, the only relevant fact for the Actor (that state B is the best one) is hidden away in the small relative differences between the states' values. Williams (1992) introduced the concept of the '*baseline*' to solve this problem. The idea behind this approach is quite simple. By subtracting the same amount from all states' values, the important information is emphasized. Therefore, in the example above, a baseline of 100 could be used, to tell the learner how advantageous state B is compared to A and C. The only constraint is that the baseline be independent from the action variable. Based on the same principle, Mnih et al. (2016) define the Advantage function as:

$$A(S_t;\omega_t) = \sum_{i=0}^{k-1} \gamma^i \times r_{t+i} + \gamma^k \times v(S_{t+k};\omega_t) - v(S_t;\omega_t) \tag{C.5}$$

Where $k$ can vary between episodes, but is upper bounded by the number of steps between two learning phases. For each time step $t$, this quantity represents the discrepancy between the expected value of a given state and an estimation of its actual value, also known as the TD error. Consequently, introducing the Advantage within the learning rule means, that rather than simply telling the agent how good an action was, it now knows just how better than expected said action actually is. The resulting expression for updating the network's weights is:

$$\theta_{t+1} = \theta_t + \varepsilon \times [R_t - v(S_t;\omega_t)] \times \left[\lambda^{\theta}\mathbf{z}_{t-1}^{\theta} + \nabla \log\left(\pi(a_t|S_t;\theta_t)\right)\right] \tag{C.6}$$

Where $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$, is defined as the accumulation of the discounted rewards expected by the agent at time $t$ and in state $S_t$. Finally, to improve exploration an entropy term $H$ is added to the learning rule. In this case, the entropy can be seen as a measure of how certain the Actor is when choosing a given action. Intuitively, if the Actor outputs similar probabilities for all possible actions, the entropy will be high and so will the corresponding weights update. On the contrary, if one action has a much higher probability of being chosen compared to the others, then the entropy will be low. Thus, resulting in a smaller change in weights. Including the entropy, the rule for updating the Actor's policy can be reformulated as:

$$\theta_{t+1} = \theta_t + \varepsilon \times [\beta \nabla H\left(\pi(a_t|S_t;\theta_t)\right) + (R_t - \nu(S_t;\omega_t)) \times \nabla \log\left(\pi(a_t|S_t;\theta_t)\right)] \tag{C.7}$$