A BDI empathic agent model based on a multidimensional cross-cultural emotion representation

Joaquin Taverner Aparicio

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Valencian Research Institute for Artificial
Intelligence (VRAIN)

A thesis submitted for the degree of
Doctor of Philosophy in Computer
Science

Supervised by
Dr. Vicente Botti
Dr. Emilio Vivancos

January 2022

Supervisors

Dr. Vicente Botti
Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València (Spain)

Dr. Emilio Vicancos
Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València (Spain)

External Reviewers

Dr. Abe Kazemzadeh
University of St. Thomas, United States

Dr. Juan Manuel Corchado
University of Salamanca, Spain

Dr. Stacy Marsella
Khoury College of Computer Sciences, United States

Thesis Defense Committee Members

Dr. Ana Paiva
University of Lisbon, Portugal

Dr. Juan Manuel Corchado
University of Salamanca, Spain

Dr. Vicente Julian Inglada
Universitat Politècnica de València, Spain

i

Abstract

Human beings are, by nature, affective beings; our emotions, moods, personality, or relationships with others guide our motivations and our decisions. One of the main cognitive abilities related to affect is empathy. Empathy is a psychological construct whose definition has evolved over the years and whose meaning refers to a wide range of affective and cognitive competencies that are fundamental in the development of human beings as social beings. The use of empathy in the field of artificial intelligence can revolutionize the way we interact with machines as well as the way we simulate human behavior.

On the other hand, it must be considered that human beings usually resort to the use of different words such as "sad" or "happy" to express or verbalize our affective state. However, these words are simplifications that cover a wide spectrum of cognitive processes and mental states. Moreover, it should be considered that these words have a high dependence on the language and culture in which they are used. Therefore, computational representation models of affective states must adaptable to different cultural environments and to allow an agent to express or represent, by means of words, a given affective state.

In this thesis, a new model of empathic agent capable of adapting its behavior to different cultural environments is proposed. To this end, first, a new experiment-based methodology to adapt an emotion representation space based on the dimensions of pleasure and arousal for simulation and affective computational recognition to different cultural environments is presented. The results of an experiment conducted with European Spanish speakers are used to provide a new fuzzy logic-based model for representing affective states in the dimensions of pleasure and arousal using a polar coordinate approach. To prove that cultural differences affect the pleasure and arousal values associated with each emotion, the experiment was replicated with Portuguese and Swedish participants. Secondly, a new model of emotion elicitation in affective agents using fuzzy logic is presented. The emotions generated in the agent by the fuzzy appraisal rules are expressed in the model of affect representation resulting from the previously described experiments. In addition, a new affect regulation process is proposed to adapt the agent's mood,

represented by a vector in the pleasure-arousal space, when an emotion is
elicited. Third, a formalization of the syntax, semantics and reasoning cycle
of AgentSpeak to enable the development of affective agents with empathic
capabilities is presented. Drawing on the theories of empathic appraisal and
empathic regulation, the agent's reasoning structure is modified to allow em-
pathy to affect the decision-making process. Finally, a model of an empathic
pedagogical agent for education on good practices in the use of social net-
works is introduced. The agent is able to recognize the user's emotion when
interacting with the social network. Based on the user's emotion and behav-
ior in the social network, the agent estimates a plan to educate the user in
the correct and secure use of social networks.

Resumen

Los seres humanos somos por naturaleza seres afectivos, las emociones, el estado de ánimo, nuestra personalidad, o nuestras relaciones con los demás guían nuestras motivaciones y nuestras decisiones. Una de las principales habilidades cognitivas relacionadas con el afecto es la empatía. La empatía es un constructo psicológico cuya definición ha ido evolucionando a lo largo de los años y cuyo significado hace referencia a un amplio abanico de competencias afectivas y cognitivas que son fundamentales en el desarrollo del ser humano como ser social. El uso de la empatía en el ámbito de la inteligencia artificial puede revolucionar la forma en la que interactuamos con las máquinas así como la forma en la que simulamos el comportamiento humano.

Por otro lado, hay que tener en cuenta que los seres humanos habitualmente acudimos al uso de distintas palabras como "triste" o "contento" para expresar o verbalizar el estado afectivo. Sin embargo, estas palabras son simplificaciones que abarcan un amplio espectro de procesos cognitivos y estados mentales. Además, hay que considerar que estas palabras tienen una alta dependencia del idioma y la cultura en la que se utilizan. Por tanto, los modelos de representación computacional de los estados afectivos deben se capaces de adaptarse a distintos entornos culturales y de permitir que un agente exprese o represente, mediante palabras, un determinado estado afectivo.

En esta tesis se propone un nuevo modelo de agente empático capaz de adaptar su comportamiento a distintos entornos culturales. Para ello, en primer lugar, se presenta una nueva metodología basada en la experimentación para adaptar un espacio de representación de emociones basado en las dimensiones del placer y la activación para la simulación y el reconocimiento computacional afectivo a diferentes entornos culturales. Los resultados del experimento realizado con hispanohablantes europeos se utilizan para proporcionar un nuevo modelo basado en la lógica difusa para representar estados afectivos en las dimensiones de placer y activación utilizando un enfoque de coordenadas polares. Para demostrar que las diferencias culturales afectan a los valores de placer y activación asociados a cada emoción, el experimento se repitió con participantes portugueses y suecos. En segundo lugar, se presenta un nuevo modelo de elicitación de emociones en agentes afectivos que utiliza

lógica difusa. Las emociones generadas en el agente por las reglas de valoración difusa se expresan en el modelo de representación del afecto resultante de los experimentos previamente descritos. Además, se propone un nuevo proceso de regulación del afecto que adapta el estado de ánimo del agente, representado mediante un vector en el espacio placer-activación, cada vez que una emoción es elicitada. En tercer lugar, se propone una formalización de la sintaxis, la semántica y el ciclo de razonamiento de AgentSpeak para permitir el desarrollo de agentes afectivos con capacidades empáticas. Partiendo de las teorías de valoración empática y regulación empática, se modifica la estructura de razonamiento del agente para permitir que la empatía afecte al proceso de toma de decisiones. Finalmente, se presenta un modelo de agente pedagógico empático para la educación sobre buenas prácticas en el uso de las redes sociales. El agente es capaz de reconocer la emoción del usuario cuando interactúa con la red social. En base a la emoción del usuario y su comportamiento en la red social, el agente estima un plan para educar al usuario en el uso correcto y seguro de las redes sociales.

Resum

Els éssers humans som per naturalesa éssers afectius, les emocions, l'estat d'ànim, la nostra personalitat o les nostres relacions amb els altres guien les nostres motivacions i les nostres decisions. Una de les habilitats cognitives principals relacionades amb l'afecte és l'empatia. L'empatia és un constructe psicològic la definició del qual ha anat evolucionant al llarg dels anys i el significat del qual fa referència a un ampli ventall de competències afectives i cognitives que són fonamentals en el desenvolupament de l'ésser humà com a ésser social. L'ús de l'empatia en l'àmbit de la intel·ligència artificial pot revolucionar la forma en la qual interactuem amb les màquines així com la forma en què simulem el comportament humà.

D'altra banda, cal tenir en compte que els éssers humans habitualment acudim a l'ús de diferents paraules com "trist" o "content" per expressar o verbalitzar l'estat afectiu. Tot i això, aquestes paraules són simplificacions que abasten un ampli espectre de processos cognitius i estats mentals. A més, cal considerar que aquestes paraules tenen una alta dependència de l'idioma i la cultura en què s'utilitzen. Per tant, els models de representació computacional dels estats afectius han de ser capaços d'adaptar-se a diferents entorns culturals i de permetre que un agent expresse o represente, mitjançant paraules, un estat afectiu determinat.

En aquesta tesi es proposa un nou model d'agent empàtic capaç d'adaptar el seu comportament a diferents entorns culturals. Per això, en primer lloc, es presenta una metodologia nova basada en l'experimentació per adaptar un espai de representació d'emocions basat en les dimensions del plaer i l'activació per a la simulació i el reconeixement computacional afectiu a diferents entorns culturals. Els resultats de l'experiment realitzat amb hispanoparlants europeus es fan servir per proporcionar un nou model basat en la lògica difusa per representar estats afectius en les dimensions de plaer i activació mitjançant un enfocament de coordenades polars. Per demostrar que les diferències culturals afecten els valors de plaer i activació associats a cada emoció, l'experiment es va repetir amb participants portuguesos i suecs. En segon lloc, es presenta un nou model d'elicitació d'emocions en agents afectius que fa servir lògica difusa. Les emocions generades a l'agent per les regles de

valoració difusa s'expressen en el model de representació de l'afecte resultant dels experiments descrits prèviament. A més, es proposa un nou procés de regulació de l'afecte que adapta l'estat d'ànim de l'agent, representat mitjançant un vector a l'espai plaer-activació, cada cop que una emoció és elicitada. En tercer lloc, es proposa una formalització de la sintaxi, semàntica i cicle de raonament d'AgentSpeak per permetre el desenvolupament d'agents afectius amb capacitats empàtiques. Partint de les teories de valoració empàtica i regulació empàtica, es modifica l'estructura de raonament de l'agent per permetre que l'empatia afecti el procés de presa de decisions. Finalment, es presenta un model d'agent pedagògic empàtic per a l'educació sobre bones pràctiques en l'ús de les xarxes socials. L'agent és capaç de reconèixer l'emoció de l'usuari quan interactua amb la xarxa social. En base a l'emoció de l'usuari i el seu comportament a la xarxa social, l'agent estima un pla per educar l'usuari en l'ús correcte i segur de les xarxes socials.

Contents

List of Figures

List of Tables

Part I

Introduction and objectives

Introduction and objectives

"The heart has its reasons which reason knows not of" – Cite associated with the French mathematician, physicist, and philosopher Bleis Pascal [162].

Today's society is increasingly moving towards the use of computer systems and virtual environments. Human-machine interaction has become an everyday phenomenon in both personal life and workplace. In recent years, a special effort has been made to develop systems capable of simulating more natural human-like interactions [40, 115]. Generally, the approaches proposed in this area are based on the agent-oriented paradigm. Software agents are able to reason, act proactively, and cooperate with other agents to establish plans that allow them to achieve their objectives [28]. Traditionally, most of the proposals using the agent-oriented paradigm are based on some types of logic and economic theory, generally through the maximization or minimization of an utility function, to carry out the agent's action planning. One of the best known model that allows the development of "rational" agents is the BDI (beliefs-desires-intentions) model that is based on Aristotelian practical reasoning. This type of practical reasoning is based on the intention to achieve a goal and a reasoned decision-making process performed to select an action (or plan of actions) from a set of possible actions to achieve that goal [225]. Over the years, several agent models capable of simulating the human "logical reasoning" process in controlled environments have been proposed. However, this type of agents based on classical practical reasoning present some limitations regarding the system's capacity to simulate certain human behaviors. This is because human behavior does not always respond to logic but is often guided by emotions and other affective, intrapersonal, or interpersonal factors. There are several studies from different fields of psychology, sociology, ethology, and neuroscience that relate diverse aspects of affect (e.g., emotions, mood, or personality) with the capacity for reasoning, judgment, decision-making, or risk tolerance (among others) [8, 80, 240]. For example, empirical evidence suggests that there is a critical impact of emotions on cognition and a high variability of this impact among individuals with different personality factors. This is because personality can make a person more or less likely to experience certain types of mood [33, 133]. Therefore, one of the most significant challenges that the research community on human simulation and human-machine interaction faces, is to succeed in simulating affective and social human behavior through credible intelligent agents or virtual characters as well as system personalization, i.e., generating systems capable of

been adapted to different user characteristics to offer a customized experience [37, 42, 241]. In order to simulate this type of "non-logical rational" behavior through intelligent agents, it is necessary to consider the factors related to social interaction, social cognition, intrapersonal or interpersonal characteristics, affective state, or social context (among others) when establishing the agent's reasoning, decision-making, or behavior selection mechanisms. However, systems capable of simulating human behavior at a social and affective level are complex and require high level abstractions and metaphors to describe capabilities and characteristics that cannot be explained by classical low level models. These metaphors and abstractions are aimed at simulating human behaviors with a higher level of realism as well as reflecting more closely the way in which humans understand the world. In this sense, over the years, in the field of affective computing [170], different models of agents capable of recognizing, processing, and simulating different abilities related to human affective behavior have been proposed [128]. These models typically include different cognitive components such as affective states or personality that come from different branches of psychology, ethology, philosophy, and sociology. Affective states, which encompass emotions and moods, are complex constructs whose definition remains the subject of debate to this day. Although the difference between emotions and mood continues to be a topic of debate [19], in general, it is accepted that an emotion is a rapid response to a given stimulus, while mood has a longer duration (from minutes to days) and a lower intensity than emotions and is not related to any particular stimulus. Instead, mood is produced by a succession of stimuli and other factors such as the context or the person's personality [64].

Over the years, several theories have been proposed in an attempt to explain how affective states are generated in the human brain. One of the most popular theories is the appraisal theory [72, 163]. According to this theory, when a stimulus is perceived, an appraisal process is produced, resulting in an emotion [135]. These theories are based on the existence of different variables known as *appraisal variables* [231], whose number and type vary depending on the author. Therefore, it is necessary to determine the mechanisms associated with these appraisal theories and how to translate these mechanisms into agent model [182].

On the other hand, there are several theories that study the representation of affective states in the human brain as well as the processes that are related to them. In this sense, there are several theories of representation of affective states that can be generally classified into two main groups: categorical and dimensional [93]. On the one hand, categorical approaches are based on the use of a limited set of words or labels to define the different affective states [66]. These theories are based on how humans express and recognize complex cognitive concepts, which represent affective states, through the use of different words such as "sadness" or "joy". This way of representation is commonly used in the area of affective computing to generate models for recognition, interpretation, processing, and simulation of affective states [17, 20, 74]. These words provide a means to verbalize affective states, which is a key factor in human affective interactions. One of the best known models is the model of basic emotions proposed by P. Ekman in which six labels are defined: *sadness*, *happiness*, *anger*, *fear*, *disgust*, and *surprise* [65]. This model is one of the most widely used in the area of affective computing to express, recognize, and represent affective states [20, 173].

In contrast, dimensional approaches allow to reduce the level of abstraction by representing affective states through different dimensions. These theories have their origin in psychological constructivism, according to which, affective states cannot be established by a finite set of labels since they depend on the individual's subjective experience that occurs in a particular context (e.g., a particular culture or language). One of the most representative dimensional models is the *Core Affect* model proposed by J.A. Russell [194]. In that model, affective experiences are defined as emotional cues and are represented by two dimensions: pleasure and arousal. Russell also evidenced through experimentation that the labels used to define the different affective states followed a circular pattern along these two dimensions. From these results he developed the *Circumplex Model of Affect* [174] that has also been used in affective computing [204, 243].

Moreover, several studies over the years have shown that emotions have a high dependence on language and culture [23, 46, 196]. The same word or label used to define an emotion or affective state can have a different conceptualization and be interpreted in different ways depending on the cultural

environment. In fact, there are words that express affective states that are very difficult to translate into other languages [107, 137]. Therefore, to establish generic models of affect representation capable of adapting to different cultural environments, it is necessary to establish methodologies that allow to adapt these models to other languages and cultures.

In recent years there is a growing trend towards the development of agents with skills related to social cognition [142, 165, 236]. Social cognition is a construct used to encompass the cognitive processes that underlie the development of social interactions and relationships, including phenomena such as social perception, Theory of Mind, and empathy. Social perception refers to the cognitive ability of evaluating the social context (e.g., norms, rules, or relationships) in order to make judgments about other individuals involved in that social context [77, 228]. Theory of Mind (also referred to as perspective-taking) is a cognitive ability that allows one to understand and predict the mental states, behaviors, beliefs, and intentions of others [226]. Theory of Mind is a fundamental component in the development of other high level cognitive and social skills such as empathy. Finally, empathy is a construct that encompasses different processes aimed at understanding and feeling the affective state of others [22, 50]. Several authors consider empathy as an inherent factor in the development of the human being as a social individual [22, 100, 184, 157, 205]. Empathy also plays a fundamental role in our society, affecting both morality and mutual understanding. [99]. In fact, empathy appears to be the basis on which social behaviors and interpersonal interactions are founded [50, 149]. Empathy connects people by promoting relationships and collaborations through the exchange of experiences, needs, and desires. [184]. However, the definition of the term empathy continues to divide the academic community [164]. This is largely due to the fact that the concept of empathy encompasses a large number of emotional, ethical, moral, and social factors. This lack of consensus can be clearly seen in the review conducted in [49], in which the authors identified forty-three different definitions for the concept of empathy.

For all the above mentioned reasons, empathy is postulated as a fundamental piece in the development of models focused on the simulation of human social behavior as well as in the improvement of human-machine interaction

[233]. In fact, systems that are capable of simulating some level of empathic abilities have been used in different contexts and have proved to be more reliable and more credible, thus reducing the stress and frustration of human users and improving human-machine interaction [165]. However, this type of models are generally presented as ad-hoc proposals for a predetermined scenario and with a limited capacity to simulate emotional, social, or empathic behavior [32, 138, 189]. Therefore, much work remains to be done in the areas of affective computing, simulation of human behavior, and human-machine interaction until a reliable simulation of empathic interactions and human behavior will be achieved.

1.1 Motivation

In recent years we have experienced an exciting evolution in the field of artificial intelligence and affective computing. However, we are still far from being able to understand and reliably simulate the cognitive processes that occur in the human brain and that constitute the foundation of human behavior. One of the remaining challenges is to generate models to simulate human social and affective behavior. Motivated by this challenge, at the beginning of this thesis we proposed the design of an affective agent model capable of simulating empathic interactions through an automatic reasoning process that would allow to adapt the agent to different cultural environments, contexts, and social situations.

One of the first difficulties that appear in this type of system is the way in which the agent's knowledge is represented. For a software agent to be able to reason using abstract concepts such as emotions, mood, or empathy, it is necessary to study how these constructs are produced and verbalized. If we pay attention to the usual way in which human beings refer to emotions or moods we observe that we generally use terms such as "very sad" or "happy". Therefore, it is necessary for the system to be able to recognize this terminology and apply it in its affective reasoning system. This can be achieved by a model of affective states representation that will allow the agent to recognize and use this terminology as well as to apply it in its affective reasoning

process.

There are several models of representation of affective states that are commonly used in the field of affective computing. Categorical models, such as Ekman's basic emotions model [66], provide a simple way of representing affective states by means of a finite set of affective words or labels. However, by a label based representation we are assuming a high level of abstraction since we are condensing the entire spectrum of affective states into a limited set of labels. Therefore, we are losing part of the affective information, such as the intensity of the affective state, or the probability of moving from one affective state to another depending on a certain stimulus [71, 163, 232]. In addition, the use of these words to define affective and cognitive terms is subject to possible variations between different cultures and languages [151, 196]. For example, the German vocabulary has an extensive catalog of affective labels defining emotions such as *Schadenfreude* which can be summarized as "pleasure from the suffering of others" or *Sehnsucht* which can be described as "a strong desire for an alternative life" but whose overall conceptualization is very difficult to describe by means of a word or label in other languages such as English or Spanish [107, 137]. Therefore, translating emotional words or labels from one language to another can lead to cultural misunderstandings as well as undesired behaviors that could harm the simulation of affective behavior and the user's experience.

On the other hand, dimensional approaches, such as the *Circumplex Model of Affect*, propose a representation in a continuous space based on a limited set of dimensions. However, the direct use of this model in affective computing, presents a number of disadvantages probably because it was not designed for its use in a computational system. The first disadvantage is evident from the way in which emotions are represented by a single point in the pleasure-arousal space. This point-based representation makes it difficult to explore the relationship between the dispersion of emotions and the levels of pleasure and arousal and is far from the way in which emotions are understood by humans, as we prove in the experiment presented in Section 2.3. Another clear disadvantage is that that model does not consider the relationship between the dimensions of pleasure and arousal and the intensity of the emotions. Finally, it is necessary to consider the cultural dependence of the model, i.e., the

model was designed from an experiment with English-speaking participants and therefore a direct application to other languages may not be appropriate.

On the other hand, in the development of models to simulate human affective behavior, it is necessary to establish mechanisms to allow the agents to simulate the suitable emotions according to the perceived stimulus or event. In that sense, one of the most relevant theories to explain the phenomenon of emotion in the field of psychology is the appraisal theory. In the literature we find several models of affective appraisal. However, most of the models used in affective computing use numerical variables, which is far from the way humans interpret cognitive states [81, 188]. Moreover, to our knowledge, current computational appraisal models do not allow the elicitation of emotions with intensity. However, the intensity of affective states is an important part of human affective behavior.

1.2 Objectives

Considering the motivations described in Section 1.1 the main objective of the research process conducted in this thesis is the development of a formal model of an affective agent capable of simulating empathic interactions (both with human users and software agents), of expressing its affective states using the same terms used by human beings, and of adapting its behavior to different cultural environments. In order to achieve this goal it is necessary to develop (i) a model for representing and expressing affective states, (ii) a methodology to adapt this representation model to different cultural environments, (iii) an appraisal process to elicit emotions in affective agents, and (iv) a formal definition of an affective agent that use the appraisal and the representation of affective states to simulate empathic interactions. Therefore, the following specific objectives are proposed:

1. Review of the state of the art.

    1.1 Analyze the theories related to affective states, affective state dimensional representation and cultural difference, emotion appraisal,

and empathy from the perspectives of psychology, neuroscience, sociology, and computating.

1.2 Analyze of the state of the art related to the development of affective agents focusing on the approaches that consider empathic behaviors.

2. Design a cross cultural computational model for affective state representation

2.1 Determine the structure of a representational model considering current theories in the fields of psychology and neuroscience.

2.2 Design of a method to discover the structure of affective states in the representation model.

2.3 Conduct an experiment with humans to define the representation model

2.4 Replicate the experiments in other cultural environments to validate the experiment-based methodology and adapt the model to other cultures and languages.

3. Define a human-like model to elicit and express affective states.

3.1 Develop a computational model of appraisal based on fuzzy logic to use human-like affective terminology.

3.2 Define how affective states and personality influence the rational processes of an agent.

3.3 Define a set of fuzzy rules to represent and express affective states using a human-like affective terminology.

4. Design a formal model for affective agents with empathic abilities.

4.1 Design the internal processes for recognizing and representing perceived affective states and to allow the emergence of empathic emotions in an affective agent

4.2 Design the processes to adapt empathic emotions to different intrapersonal and interpersonal characteristics to simulate the empathic regulation.

4.3 Integrate this model in the general-purpose architecture *GenIA³*.

5. Propose a case of study to validate the formal design of the effective agent with empathic abilities.

1.3   Structure of the thesis

In this part we have provided an introduction to the work proposed in this thesis. Starting from the motivations, an overview of the challenges of the thesis have been provided. Then, the objectives of the research have been enumerated. The rest of the document is organized as follows:

- **Part II Contribution:** In the second part of this document the most representative contributions supporting the objectives of this thesis are presented. Following, a brief summary of each chapter is provided:

  Chapter 2 (*Selected paper*) proposes an experiment-based methodology for adapting a pleasure and arousal representation space for affective computational simulation and recognition to different cultural environments. In addition, the results of an experiment conducted with European Spanish-speakers are presented. These results are then used to provide a new cross-cultural model based on fuzzy logic to represent affective states in the pleasure and arousal dimensions using a polar coordinate approach.

  Chapter 3 shows the results of an experiment conducted in collaboration with the ALGORITMI Centre, University of Minho (Portugal) and the Department of Computing Science, Umeå University (Sweden) to adapt the computational representation model presented in **Chapter 2** to Portuguese and Swedish. The results are

compared with the experiment conducted with European Spanish-speakers, presented in **Chapter 2**, to prove that cultural differences affect the pleasure and arousal values associated with each emotion.

Chapter 4 (*Selected paper*) presents a new fuzzy appraisal model for eliciting emotions in affective agents. The emotions generated in the agent by the fuzzy appraisal rules are expressed in a pleasure an arousal representation based on the work presented in **Chapters 2** and **3**. Moreover, we present an affect regulation process that adapts the agent's mood, represented by a vector in the pleasure-arousal space, each time an emotion is elicited.

Chapter 5 describes a formalization of the syntax, the semantics, and the reasoning cycle of AgentSpeak to allow the development of affective agents with empathic abilities. Based on the theories of empathic appraisal and empathic regulation, this formalization departs from the General-purpose Architecture for Affective Agents *GenIA³* and modifies the agent's reasoning process to allow empathy to affect its decision-making process. An empathic and an affective appraisal process, based on the proposal made in **Chapter 4**, allows the agent's to elicit empathic emotions when perceiving a situation in other agents. The empthic regulation process adapts the elicited empathic emotion to different intrapersonal and interpersonal factors. The agent is also able to maintain the distinction between the self and the others.

Chapter 6 (*Selected paper*) introduces an application of the empathic agent model proposed in **Chapter 5**. This application is based on an empathic pedagogical agent model for education on good practices in the use of social networks. The agent is able to recognize the user emotions when interacting with the social network. Based on the user emotions and his/her behavior on the social network, the agent estimates the best plan to teach the user in the correct and secure use of social networks.

- **Part III Discussion:** In this part a discussions of the results obtained during the research period of this thesis is presented.

- **Part IV Conclusions and Future Work:** Finally, this last part shows the conclusions of the research conducted in this thesis and propose possible future works.

1.4 List of publications

The work done in this thesis is supported through the following publications (note that publications marked with * are those included in this thesis document):

- Journals listed in JCR:

  - *Joaquin Taverner, Emilio Vivancos, and Vicente Botti. A multi-dimensional culturally adapted representation of emotions for affective computational simulation and recognition. *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2020.3030586. *Impact Factor: 10.506.*

  - *Joaquin Taverner, Emilio Vivancos, and Vicente Botti. A fuzzy appraisal model for affective agents adapted to cultural environments using the pleasure and arousal dimensions. *Information Sciences*, 546:74–86, 2021, doi: 10.1016/j.ins.2020.08.006. *Impact Factor: 6.795.*

  - Bexy Alfonso, Joaquin Taverner, Emilio Vivancos, and Vicente Botti. From Affect Theoretical Foundations to Computational Models of Intelligent Affective Agents. *Applied Sciences*, no. 22: 10874, 2021, doi: 10.3390/app112210874. *Impact Factor: 2.679.*

- International conferences listed in CORE:

  - Joaquin Taverner, Bexy Alfonso, Emilio Vivancos, and Vicente Botti. Modeling personality in the affective agent architecture Ge-

nIA3. In *Int. Conf. on Agents and Artificial Intelligence*, pages 236–243, 2018. Doi: 10.5220/0006597202360243.

– Joaquin Taverner, Emilio Vivancos, and Vicente Botti. Towards a computational approach to emotion elicitation in affective agents. In *International Conference on Agents and Artificial Intelligence*, pages 275–280, 2019. Doi: 10.5220/0007579302750280.

– Jose Manuel Fuentes, Joaquin Taverner, Jaime Andres Rincon, and Vicente Botti. Towards a classifier to recognize emotions using voice to improve recommendations. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 218–225. Springer, 2020. Doi: 10.1007/978-3-030-51999-5_18.

– *Joaquin Taverner, Emilio Vivancos, and Vicente Botti. Towards a model of empathic pedagogical agent for educating children and teenagers on good practices in the use of social networks. In *Int. Conf. on Agents and Artificial Intelligence*, pages 439–445, 2021. Doi: 10.5220/0010345504390445.

– Andreas Brännström, Timotheus Kampik, Ramon Ruiz-Dolz, and Joaquin Taverner. A Formal Framework for Designing Boundedly Rational Agents. In *Int. Conf. on Agents and Artificial Intelligence*, (Accepted), 2022.

• Other international conferences:

– Joaquin Taverner, Bexy Alfonso, Emilio Vivancos, and Vicente Botti. Using different agent personalities to obtain different agent behaviours. *In Proc. of the Workshop on Affective Computing and Context Awareness in Ambient Intelligence*, 2018.

– Joaquin Taverner, Ramon Ruiz, Elena del Val, Carlos Diez, and Jose Alemany. Image analysis for privacy assessment in social networks. In *International Symposium on Distributed Computing and Artificial Intelligence*, pages 1–4. Springer, 2018. Doi: 10.1007/978-3-030-00524-5_1.

– Joaquin Taverner, Emilio Vivancos, Vicente Botti, and Bexy Alfonso. Influence of the agent personality on its mood. In *Proc. of the Workshop on Affective Computing and Context Awareness in Ambient Intelligence*, 2019.

1.5   List of projects

- FPI grant

  – *Funder:* Generalitat Valenciana and European Social Fund (ACIF/ 2017/085).
  – *Lead Applicant:* Joaquin Taverner Aparicio.
  – *Years:* 2017 - 2021.

- Intelligent Agents for Privacy Advice in Social Networks (AI4PRI).

  – *Funder:* Ministerio de Economia y Empresa (TIN2017-89156-R).
  – *Lead Applicant:* E. Argente, and A. García-Fornes.
  – *Years:* 2018 - 2021.

- Affective Intelligent Agents for Persuading Civic Behavior in Virtual Environments

  – *Funder:* Ministerio de Economia y Empresa (PID2020-113416RB-I00).
  – *Lead Applicant:* E. Argente, and A. García-Fornes.
  – *Years:* 2021 - 2024.

Part II

Contribution

A multidimensional culturally adapted representation of emotions for affective computational simulation and recognition

JOAQUIN TAVERNER, EMILIO VIVANCOS, AND VICENTE BOTTI

IEEE Transactions on Affective Computing

Doi: https://doi.org/10.1109/TAFFC.2020.3030586

Abstract

One of the main challenges in affective computing is the development of models to represent the information that is inherent to emotions. It is necessary to consider that the terms used by humans to name emotions depend on the culture and language used. This paper presents an experiment-based method to represent and adapt emotion terms to different cultural environments. We propose using circular boxplots to analyze the distribution of emotions in the Pleasure-Arousal space. From the results of this analysis, we define a new cross-cultural representation model of emotions in which each emotion term is assigned to an area in the Pleasure-Arousal space. An emotion is represented by a vector in which the direction indicates the type, and the module indicates the intensity of the emotion. We propose two methods based on fuzzy logic to represent and express emotions: the emotion representation process in which the term associated with the recognized emotion is defuzzified and projected as a vector in the Pleasure-Arousal space; and the emotion expression process in which a fuzzification of the vector is produced, generating a fuzzy emotion term that is adapted to the culture and language in which the emotion will be used.

2.1 Introduction

Emotions play an important role in our behavior as humans. Emotions are present in our day-to-day relationships with other human beings and when we face any situation or event. Systems that are capable of recognizing, processing, or simulating emotions can improve the interaction between humans and machines by making this interaction more natural and realistic.

Over the years, different models have been proposed to recognize emotions in human beings. With the improvement of machine learning, there is currently a large number of models that recognize emotions using images, voice, text, or electroencephalograms. However, most of these models focus on improving the accuracy of the recognition without paying much attention to the way of representing that knowledge for its possible use in systems such as affective agents. An affective agent is a system that is capable of simulating human affective behaviour. Currently, there are several approaches for affective agents in which different models of representation of emotions are used [6, 81, 148, 155, 182]. Frequently, proposals made from affective computing represent emotions using simple labels (e.g., *Happy* or *Sad*). However, a representation of emotions that uses a continuous multidimensional space seems to be more appropriate for use in computational models since it allows storing more information about emotions, such as the intensity of emotions or the proximity between emotions [34]. This representation also provides a greater capacity to analyze the variations that occur in emotions and mood over time, thereby improving the simulation of human emotional behavior.

Different models of representation of emotions have been proposed from psychology. One of the best known is *The Circumplex Model of Affect* proposed by J. A. Russell [193] (Fig. 2.1). This model relates emotions with their *Pleasure* and *Arousal* values. In addition, according to several studies, emotions greatly depend on language and culture [196, 23, 46]. In other words, the same emotion label can be interpreted in a different way depending on the culture and the language. Nevertheless, even though cultural and language factors are taken into account in other domains such as recommendation systems, to our knowledge, there are still no proposals in the affective computing area that really consider cultural and language factors to represent

and express emotions in a computational system.

On the other hand, there is a recent tendency to use emotional models based on fuzzy logic [150, 13]. This is due to the fact that this type of knowledge representation is closer to the way in which human beings express their emotions using terms such as *very happy* or *a little happy* [122]. Therefore, a model that represents human emotions should be able to deal with this type of fuzzy terminology.

This paper proposes an experiment-based method that represents emotions in a *Pleasure-Arousal* dimensional space adapted to the language and culture where the emotions will be expressed. This model uses fuzzy logic to better simulate the way in which humans express emotions. The rest of this paper is organized as follow. In Section 2.2, we discuss the most important proposals for emotion representation in affective computing. Section 2.3 introduces an experiment-based method to adapt the emotion representation to the cultural environment in which emotions are used. The fuzzy culturally adapted model of affect obtained from the experiment is presented in Section 2.4. Finally, the main conclusions and some future works are presented in Section 2.5.

2.2 Related works

The term emotion is a psychological construct whose definition has long been the subject of debate. Over the years, different theories have been proposed to explain emotional behavior. Some theories, such as basic emotion theories, propose that there is a limited number of emotions and that these emotions are understood universally. One of the best known theories in this area is the Basic Emotion Theory proposed by Ekman [65]. Ekman's theory is based on six basic emotions: *Happiness*, *Surprise*, *Fear*, *Anger*, *Disgust*, and *Sadness*. Plutchik defined his theory using eight basic emotions [172] in which different feelings can be derived from the combination of two adjacent emotions (Fig. 2.2). According to these theories, one event is universally related to a single emotion.

In contrast, the constructivist theories contradict the basic emotions theories stating that emotion labels/words do not have a universal meaning.
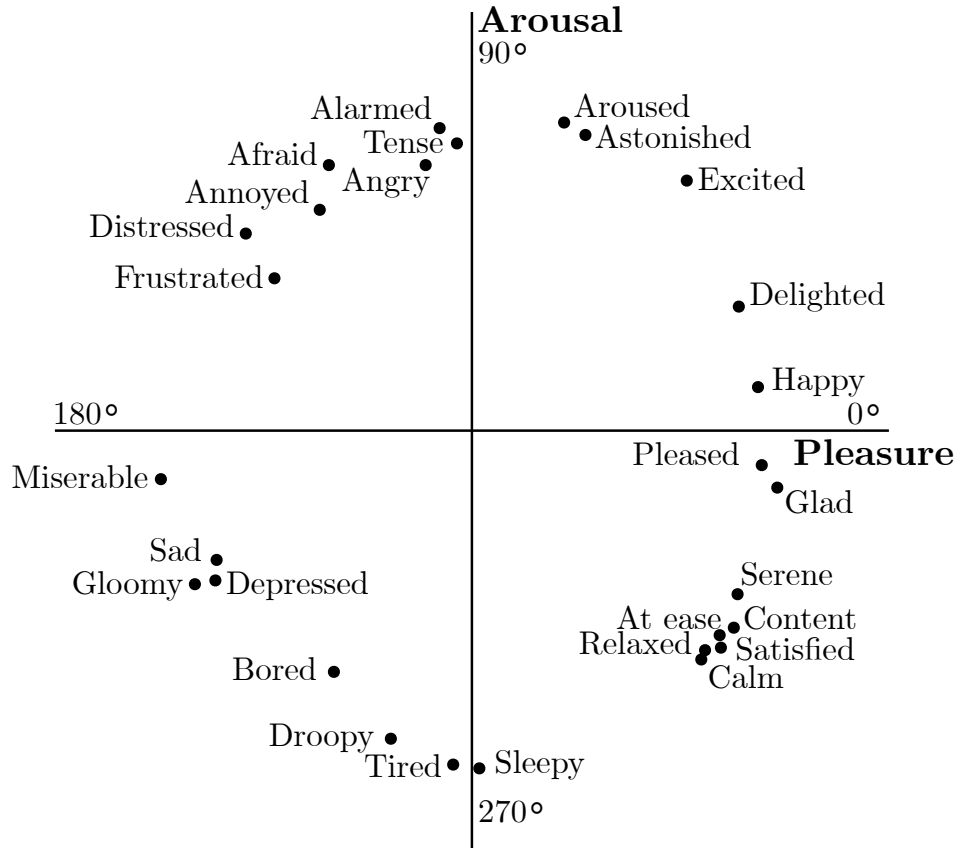
Figure 2.1: The Circumplex Model of Affect

According to constructivism, the labels that we use to refer to emotions depend on the culture and the language used. Therefore, there is no direct correspondence for all of the emotion labels among different languages [97]. For example, in the German language, there is an emotion called *Schadenfreude*, whose meaning could be translated into English as "pleasure from the suffering of others" but this translation does not quite capture the complete meaning of the emotion as it is shown in [137].

One of the most cited constructivist theorists is J. A. Russell. In his *Core-Affect* theory [196] language and culture have an important weight when cataloging different emotions. He also proposed that emotions are related to the dimensions of *Pleasure* (sometimes called *Valence*) and *Arousal* (sometimes called *Activation*) [194]. He observed through experimentation that emotions follow a circular pattern within a continuous space based on these
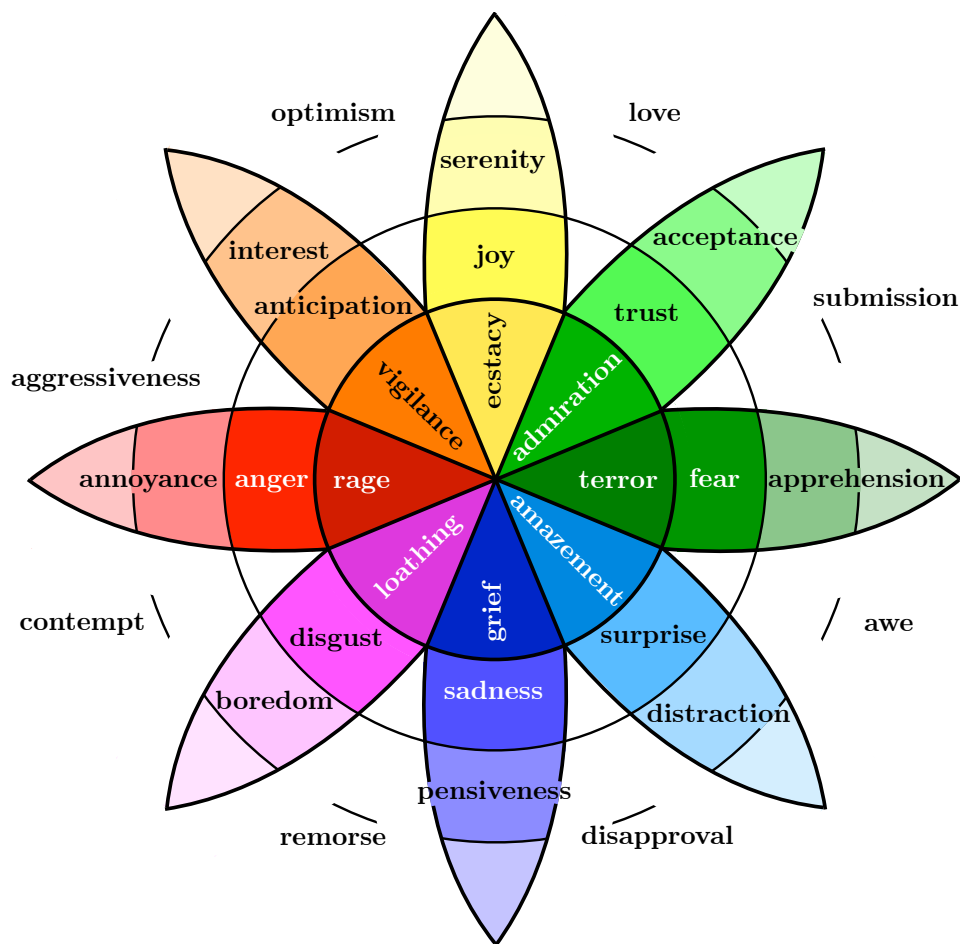
Figure 2.2: Wheel of emotions proposed by Plutchik [172]

two dimensions. He called this pattern *The Circumplex Model Of Affect* [193] (Fig. 2.1), which shows the relationship between emotions and the values of *Pleasure* and *Arousal*.

On the other hand, Reisenzein [183] observed that the intensity of emotions is proportionally related to the values of *Pleasure* and *Arousal*: the greater the value of these dimensions, the greater the intensity of the emotion. For example, the intensity of the *Happy* emotion is more related to the level of *Pleasure*, while the intensity of the *Alert* emotion is more related to the level of *Arousal*. In addition, he also observed that low levels of *Pleasure* and *Arousal* mean that the intensity of emotions is so low that it can be assumed that there is no emotion.

Mehrabian proposed adding a third dimension, known as *Dominance*, to the *Pleasure-Arousal* space. This created a new model known as the PAD model [154], which is one of the most widely used models when designing affective agents that simulate emotional behavior. The *Dominance* dimension increases the space of representation allowing some emotions to be disambiguated. However, many models propose that this third dimension must be considered as an appraisal variable rather than as a representation variable [147].

Jackson et al. [107] analysed around 2500 languages to determine the level of similarity of 24 emotion labels across different cultures. They studied six dimensions associated with emotions (*Pleasure*, *Arousal*, *Dominance*, *Certainty*, *Approach-avoidance*, and *Sociality*) and detected that *Pleasure* and *Arousal* are the two most important dimensions to determine emotion semantics across different languages. In fact they concluded that *Pleasure* and *Arousal* are "common psychophysiological dimensions shared by all humans". Other studies showing that the behavior of emotions in the *Pleasure-Arousal* space depends on culture and language have been conducted over the years [196, 237]. Based on these findings, different psychology researchers have adapted some psychological models to different cultures and languages [57].

Emotions in Affective Computing

Currently, most of the proposals made in the affective computing area, both in the field of recognition and simulation, use Ekman's emotional model to define emotions [242, 81]. However, there are other models that use a dimensional representation of emotions [146, 13, 216]. For instance, the World Wide Web Consortium (W3C) recommends a markup language that allows both dimensional and label-based representations of emotions (EmotionML [203]).

Generally, the proposals performed in the field of affective behavior simulation are focused on the use of intelligent multi-agent systems [185, 199]. Most of these proposals take advantage of a dimensional representation to model different affective processes, such as the mood decay rate or the displacement of the mood when receiving an emotion, which would be impossible

to simulate using a label-based representation. For example, the model proposed by Gebhard in ALMA (A Layered Model of Affect) [81] uses the PAD model to represent the mood in an affective intelligent agent. On the other hand, *GenIA³* (a general purpose architecture for affective agentes) [6] is a BDI architecture created to simulate affective behavior in multi-agent systems. *GenIA³* offers a set of tools to represent affective traits like personality, emotions, and mood. This architecture also uses a dimensional representation for mood and emotions. However, these models assume that a basic emotion can be universally represented by a simple point in the dimensional space. The work proposed in this paper extends the *GenIA³* architecture through a new emotional representation model to adapt *GenIA³* affective agents to different cultures and languages.

In recent years, there seems to be a tendency to define affective models using fuzzy logic. Through fuzzy logic, it is possible to approximate the artificial representation of emotions to the way in which humans express them. For example, Jain et al. propose EMIA [108] which uses fuzzy rules to handle uncertainty in its appraisal process for five of Ekman's basic emotions. Another interesting proposal is the one presented in [121] in which a methodology to translate emotional labels from Spanish to English through fuzzy logic is proposed. The author designed an experiment in which 8 Spanish-speakers answered a questionnaire to obtain the values of *Pleasure*, *Arousal*, and *Dominance* associated with 30 Spanish emotional labels/words. From the results of the experiment, the fuzzy membership functions for each emotion were obtained. Then a comparison with the functions obtained with English-speakers in a previous experiment was performed to translate the emotional labels/words.

As the *Circumplex Model of Affect* shows, emotions follow a circular pattern in the *Pleasure-Arousal* space. This property suggests that emotions could be represented by the use of vectors in a polar coordinate system. For example, in [106], a vector based representation model of emotions using *Pleasure* and *Arousal* dimensions is proposed. The authors modify the representation space by displacing the origin of coordinates to the negative extreme of the *Arousal* dimension. However, when displacing the origin of coordinates, both the direction and the modulus of the vector are modified, which

can lead to difficulties when calculating the intensity or type of the emotion. In addition, the concept of negative *Arousal* (usually referred to as *Sleepiness* or *Deactivation*) becomes blurred, producing a decontextualization of emotions that depend on this dimension, such as *Boredom* or *Calm*. These disadvantages can be solved by the use of circular statistics.

When using circular representation spaces, such as the *Circumplex Model of Affect*, traditional statistics that are focused on linear data may not be adequate since the representation space is finite and constrained between 0 and 2π. In a circular representation space based on *Pleasure* and *Arousal*, the dispersion can be estimated by the circular standard deviation σ expressed as follows [169]:

$$\sigma = \sqrt{-2 \ \ln \overline{R}} \tag{2.1}$$

where

$$\overline{R} = \sqrt{\overline{P}^2 + \overline{A}^2} \tag{2.2}$$

where $\overline{P}$ and $\overline{A}$ represent the mean values for the *Pleasure* and *Arousal* dimensions. The distribution of the data in a circular area delimited by 2π also makes it difficult to use the probability distributions that are commonly used in linear statistics. Instead, the von Mises distribution [102] could be used to estimate the normal (Gaussian) distribution for circular data. For example, in [92], a probabilistic model based on a mixture of von Mises distributions for the recognition of emotions in faces is proposed. The authors use a representation space that is based on the dimensions of *Pleasure* and *Arousal*. The face images were projected to coordinates in the representation space to determine the type of emotion using the von Mises distribution.

Discussion

Despite the great amount of effort made to recognize and simulate emotions, there is still much to be done until we are able to understand and correctly simulate affective processes. Nowadays, there is no consensus about what is

the best way to represent emotions. This is because emotions are complex
concepts and their generation depends on multiple personal, contextual, and
cultural factors. However, in order to make better use of the current sys-
tems for emotion recognition and affect simulation, it is necessary to design
a computational model of affect that represents this knowledge as flexibly
and accurately as possible. A common representation using a continuous di-
mensional space for both recognized and simulated emotions can improve the
understanding and simulation of affective processes such as emotion elicita-
tion, emotional contagion, or empathy.

However, in affective computing, the use of emotions is usually simplified
by using categorical approaches that reduce emotions to simple labels (e.g.,
using the six basic emotions of Ekman). With this type of representation, part
of the information inherent to the emotion, such as intensity or proximity to
another emotion, is lost. Therefore, representations of emotional knowledge
based on a multidimensional continuous space are more appropriate for use in
affective computing. In a multidimensional space, emotions can move through
continuous variables representing a greater amount of information (e.g., the
intensity of the emotion) than if only labels were used.

When a dimensional representation is used, different alternatives for repre-
senting the emotions within the space arise. The most commonly used model
consists of representing each basic emotion as a point in the representation
space as is done in *The Circumplex Model of Affect* (Fig. 2.1). The main
problem that appears when using this model is the way basic emotions are
represented using a simple point in the space. This representation is far from
the way in which emotions are associated with *Pleasure* and *Arousal* by hu-
mans, as we prove in our experiment. The second problem when using this
scheme is the ambiguity that appears in the spaces between adjacent emo-
tions, because the area around each point is not clearly defined. In addition,
it is very difficult to determine categorically and without error that a point in
space corresponds to a certain emotion. Moreover, this model does not rep-
resent important information such as the intensity of the emotion. Therefore,
a more appropriate method for representing the emotions consists of defining
each emotion as an area in the emotion space, instead of a simple point as
proposed by other approaches. An area-based method better represents the

complexity and similarities among multiple emotions.

Some findings suggest that languages around the world primarily differentiate emotions on the basis of *Pleasure* and *Arousal* [107]. Therefore, when defining a computational model of affect, the culture and the language in which it is going to be used must be considered since different cultures may attribute different *Pleasure* and *Arousal* values for the same emotion label [46, 141]. For example, in the experiment presented in [196], the authors note, among other results, that the Polish participants related the calm emotion with a significantly higher level of arousal than the Greeks. Extrapolating an emotional model from one language to another without adapting it to the culture and language may produce inconsistencies, which can lead to erratic agent behavior and emotion expression. Moreover, the way in which emotions are expressed when the system interacts with humans must also be taken into account. Fuzzy logic allows affective agents to express their emotions in a similar way to humans.

2.3   An experiment-based method to adapt affective models to culture and language

Labels that are used to represent emotions do not have the same meaning in different languages and cultures, but a model that is based on the *Pleasure* and *Arousal* dimensions can represent the universal meaning of the emotions [107]. Nevertheless, the scheme proposed by Russell is a basic scheme that is used to support the circular representation of emotions. Therefore, the direct use of this scheme in a computational model can easily be criticized since it was not defined for that purpose. In addition, this scheme was generated from experiments with English-speakers. Therefore, following the foundations of constructivist theories, this representation of basic emotions may not be valid in other cultures or languages. It is necessary to design processes to adapt this emotional model to the culture and language of the environment in which the model will be used and also to discover the areas of the multidimensional space where emotions have a higher probability of occurrence. We present an

experiment-based method to find the regions representing each emotion, that can be used for different languages and cultures.

We start from the main hypothesis that Russell's model might not be valid for representing emotions in a context with a language and a culture that is different from the context in which Russell carried out his experiment. We propose a methodology that consists of two experiments. The first experiment finds the labels that best represent the meaning of the *Pleasure* and *Arousal* dimensions in the target language. In our experiment the target language is European Spanish. The second experiment finds the intervals of *Pleasure* and *Arousal* that are associated to each emotion in the target language.

For our experiment, we have selected the ten emotions used by Russell in [196]: *Fear*, *Anger*, *Disgust*, *Sadness*, *Boredom*, *Sleepiness*, *Calm*, *Happiness*, *Excitement*, and *Surprise*. The most common literal translations of these emotions into Spanish are *"Miedo"*, *"Enfado"*, *"Asco"*, *"Tristeza"*,*"Aburrimiento"*, *"Somnolencia"*, *"Calma"*, *"Felicidad"*, *"Emoción"*, and *"Sorpresa"*. However, these are not the only possible translations, and, depending on the selected word, we could be expressing different emotional meanings. Therefore, making a literal translation of emotions could lead to significant differences in the interpretation of the levels of *Pleasure* and *Arousal* of these emotions.

Method

Experiment 1: Naming the Pleasure and Arousal dimensions

Starting from the hypothesis that the literal translation of the labels associated with the *Pleasure* and *Arousal* affective categories may introduce a bias, the main objective of this first experiment is to find the words that best represent the meaning of theses affective categories in the target language. Generally, these dimensions are defined using four affective categories: *Pleasure* and its opposite *Misery*, and *Arousal* and its opposite *Sleepiness*. However, despite the fact that emotional terms are often equated in translation dictionaries, a literal translation may not accurately reflect the meaning of the original emotional term since the terms used to define emotions vary

in meaning depending on the culture and the language [23, 46]. To translate
these words we have used the following methodology:

- We consulted the translations provided by the four most widely used
  bilingual dictionaries: The Cambridge [58], Collins [59], Oxford [60]
  and WordReference [61] English-Spanish Dictionaries. We selected 30
  Spanish words for the four affective categories: 8 words for *Pleasure*, 10
  words for *Misery*, 5 words for *Arousal*, and 7 words for *Sleepiness*.

- Fifty people participated in this first experiment: 20 females and 30
  males ranging in age between 18 and 60 years old and with different
  study levels (from secondary school to the doctoral level). The instruc-
  tions were given to each participant at the beginning of the experiment.
  Each participant was asked to read each group of words and identify the
  general affective category represented by the group. Each participant
  was given an unlimited time to select the word that best represents the
  concept of the group and to remove those words that clearly do not
  express the meaning of the affective category of the group.

Experiment 2: Assigning areas to each emotion

Based on the hypothesis that the *Pleasure* and *Arousal* levels of an emotion
can vary depending on the culture and language, this second experiment
was designed to define the levels corresponding to each emotion in the target
language. We decided not to consider other dimensions because recent studies,
like [107], have shown that the *Pleasure* and *Arousal* dimensions predict the
structure in emotion across language families. Moreover, some authors have
found that human beings have difficulty and display confusion when they try
to assign a *Dominance* value to the emotions that they are feeling and that
this dimension does not show much difference between emotions [107].

One hundred people participated in this second experiment: 40 females
and 60 males ranging in age between 18 and 60 years old with different study
levels (from secondary school to the doctoral level). To avoid bias, the partic-
ipants performed the experiment individually and without a time limit. We
designed a questionnaire using the four affective categories translated into

Spanish in the first experiment. It was composed of 10 prompts, one for each emotion. At the beginning of the experiment, the instructions were given to each participant individually. For each prompt, the participants were asked to assign a level of *Pleasure* and a level of *Arousal* to the 10 selected emotions. The structure of the prompts was designed to collect the *Pleasure* and *Arousal* variables individually. To assign these two levels, each prompt was composed of two 7-item Likert scales. The scale for *Pleasure* was defined from *very miserable* to *very pleased* and the scale for *Arousal* was defined from *very sleepy* to *very aroused*. Note that the value 4 corresponds to a neutral value on these scales.

Analysis and Results

In the first experiment, we define the *degree of acceptance* of one word w for the affective category c ($accept_c(w)$) as the difference between the number of times the word w has been selected as being representative of the affective category c and the number of times the word has been removed. The *degree of acceptance* for all of the candidate words are represented in Fig. 2.3. For instance, Fig. 2.3a shows that the word *Alertado* has a *degree of acceptance* for the *Arousal* category $accept_{arousal}(alertado) = 25$, while *Excitado* has $accept_{arousal}(excitado) = -22$. The negative *degree of acceptance* means that the participants mostly rejected that word to represent the affective category of the group. The results confirm the hypothesis of the first experiment, and the need to carry out studies of this type when translating affective concepts to other languages and cultures. For example, according to the selected dictionaries, the closest translation into Spanish of the affective concept *Misery* is *Miseria*. However, as Fig. 2.3d shows, this word only has an $accept_{misery}(miseria) = -11$. Therefore, it was not selected to represent this affective category since a great number of people discarded it. For the concept of misery, the participants selected words such as *Suffering, Sadness*, and *Unhappiness* (in Spanish *Sufrimiento, Tristeza*, and *Infelicidad*). This is probably due to the fact that the concept of misery in Spanish is related more to economic poverty than to suffering. We can also see that the best translation for *Arousal* is *Alerta*. For *Sleepiness*, the words selected

(a) Arousal

(b) Sleepiness

(c) Pleasure

(d) Misery
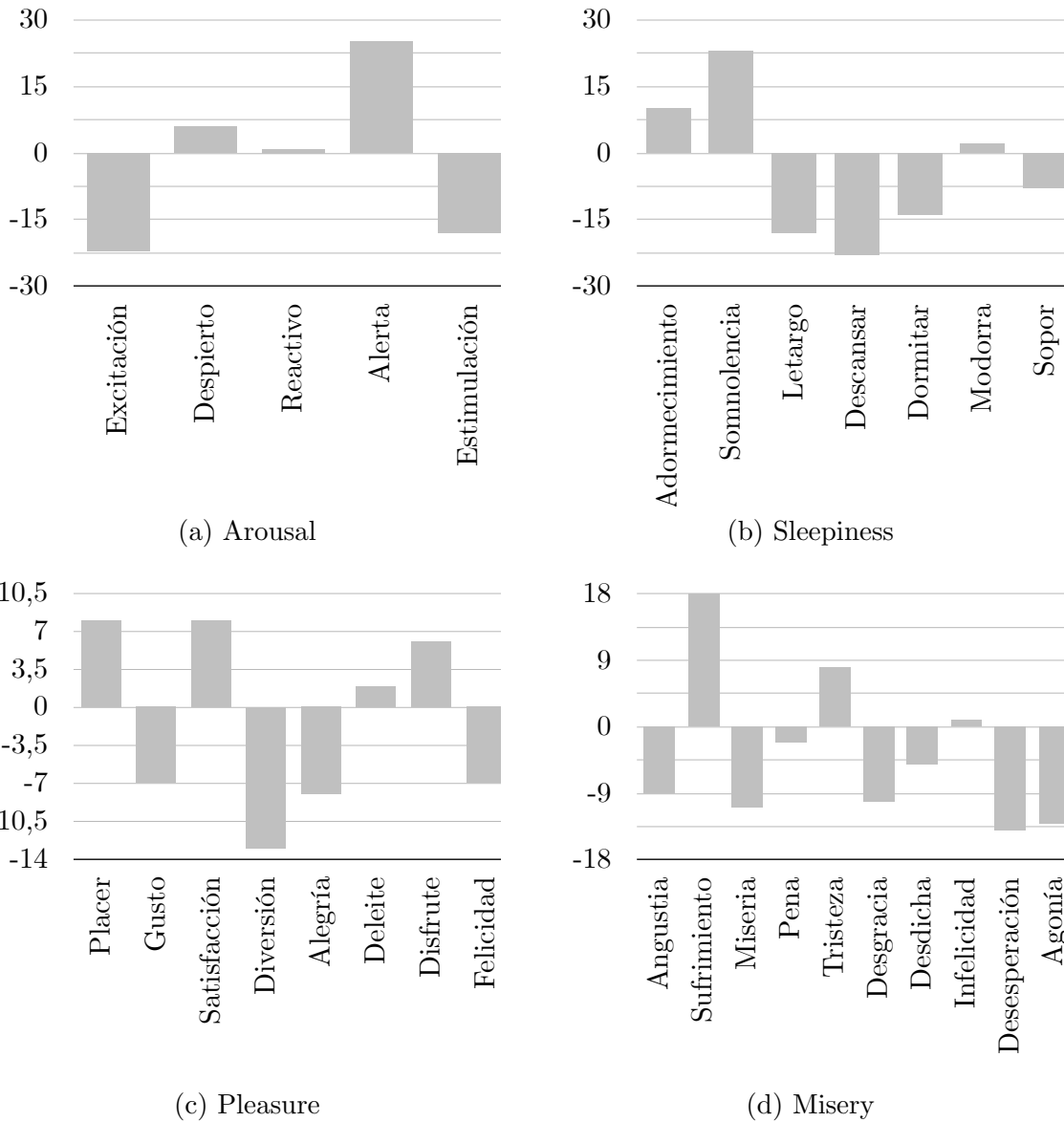
Figure 2.3: Degree of acceptance of the candidate words for each affective category.

as most representative were *Somnolencia* and *Adormecimiento*. Finally, for *Pleasure*, there is a tie between *Placer* $accept_{pleasure}(placer) = 8$ and *Satis-facción* $accept_{pleasure}(satisfaccion) = 8$. The results of the first experiment show that there are some affective categories in which there are more than one word with a positive *degree of acceptance*. For example, in the case of *Pleasure*, it is very clear that both *Placer* and *Satisfacción* are relevant to the understanding of this abstract concept. Therefore, by selecting one of the two, we could be biasing the translation of *Pleasure* since each person can have an individual conceptualization that could be influenced by their geographical location, age, or cultural environment. The conclusion we drew from these results was that, in order to avoid a possible bias that could produce misunderstandings, we needed to contextualize each affective category by using more than one word. To determine the appropriate number of words for each affective category, we used the results of the experiment. For the affective categories *Arousal*, *Sleepiness*, and *Misery*, we selected those words with a positive *degree of acceptance*, obtaining three words for each affective category. However, by applying the same criteria for *Pleasure*, four possible words were obtained. Therefore, in order to define all of the affective categories homogeneously to avoid misunderstandings, for *Pleasure*, we decided to select the three words with the highest *degree of acceptance*.

We used the second experiment to define the levels of *Pleasure*, *Misery*, *Arousal*, and *Sleepiness* associated to each emotion label. Then, we performed a circular statistical analysis of the results of the second experiment to determine the best circular representation of the selected emotions based on their *Pleasure* and *Arousal* components. Henceforth, we will refer only to the dimensions of *Pleasure* and *Arousal*, considering that *Misery* is negative *Pleasure* and *Sleepiness* is negative *Arousal*. First of all, we discarded all of the answers of the participants that represent unexpressed emotion. These emotions are characterized by zero values in the *Pleasure* and *Arousal* dimensions, which correspond to the 4$^{\text{th}}$ item in the Likert scales. We made this decision taking into account that, with low levels of *Pleasure* and *Arousal*, the intensity of the emotions is so low that it can be considered that there is no emotion [183]. Second, for all of the answers of each participant for each selected emotion, we calculated the angle α between the *Pleasure* axis and
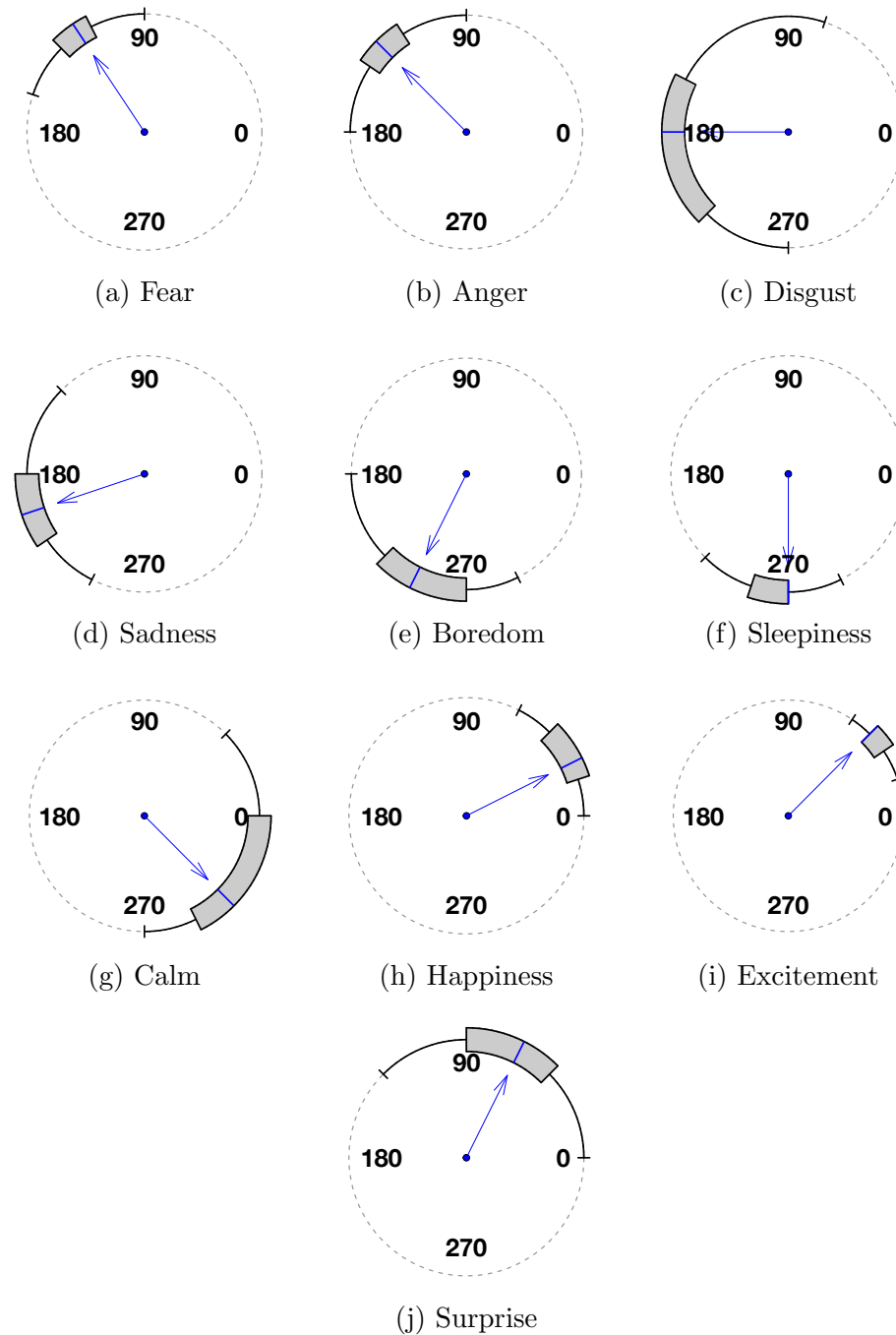
Figure 2.4: Circular boxplot graphic for the ten selected emotions in Spanish
(see Table 2.1) where the horizontal axis represents *Pleasure* and the vertical
axis represents *Arousal*.

Table 2.1: Results of the second experiment. $\overline{\alpha}_e$ and σ_e expressed in radians.

Emotion e in Spanish (English)	Mean $\overline{P}_e$	Mean $\overline{A}_e$	Mean $\overline{\alpha}_e$	Sd σ_e
Felicidad (Happiness)	0.90	0.42	0.44	0.33
Emoción (Excitement)	0.76	0.64	0.70	0.18
Miedo (Fear)	-0.58	0.81	2.19	0.27
Enfado (Anger)	-0.74	0.66	2.42	0.29
Tristeza (Sadness)	-0.96	-0.27	3.42	0.39
Aburrimiento (Boredom)	-0.41	-0.91	4.28	0.38
Somnolencia (Sleepiness)	-0.11	-0.99	4.60	0.27
Calma (Calm)	0.74	-0.67	5.55	0.63
Asco (Disgust)	-0.99	-0.04	3.19	0.76
Sorpresa (Surprise)	0.31	0.95	1.25	0.46

the line defined by the (*Pleasure*, *Arousal*) point and the origin of the coordinate system. This angle α represents the "meaning" that each participant associates to each emotion in this two-dimensional space. Third, we carried out a circular statistical analysis to eliminate possible outliers. To do this, we estimated the circular dispersion of the data associated to each emotion by the participants.

Table 2.1 summarizes the results of the second experiment: the mean values associated with the *Pleasure* and *Arousal* variables for the emotion e ($\overline{P}_e$ and $\overline{A}_e$), and the mean angle $\overline{\alpha}_e$ and circular standard deviation σ_e calculated using the Formula 2.1. To analyze the distribution of emotions in this polar representation, we use circular boxplots [36]. The circular boxplots in Fig. 2.4 show the circular distribution in the *Pleasure-Arousal* space that we obtained for the 10 emotions. As can be observed, *Sadness* is related to low levels of *Pleasure* and mean levels of *Arousal* (Fig. 2.4d), while *Excitement* is related to high levels of *Pleasure* and *Arousal* (Fig. 2.4i). These graphs also show the great differences between the dispersion levels of the emotions. For example, the dispersion of the *Excitement* emotion (Fig. 2.4i) is clearly lower than that of the *Calm* emotion (Fig. 2.4g). This confirms that, by representing the emotion with a simple point, as proposed in previous approaches, it is

Figure 2.5: Dispersion of the *Disgust* emotion.
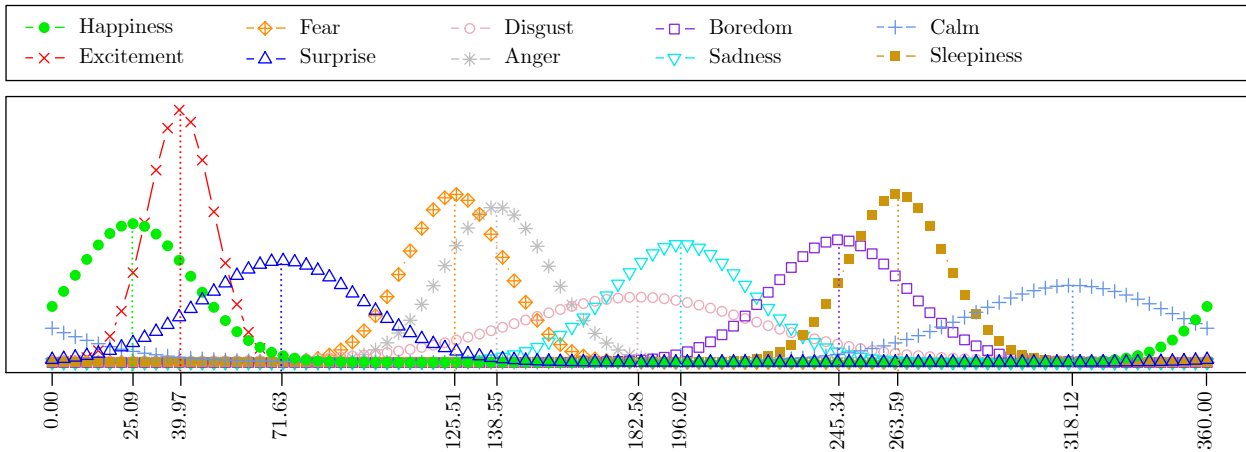
Figure 2.6: Von Mises curves of the ten Spanish emotions (see Table 2.1). The axis represents the angles α_e in degrees. Note that the order of the means of the emotions on the plot corresponds to the order of the legend.

performed a simplification that does not correspond to the area that humans assign to emotions in their minds. It also confirms the hypothesis of the second experiment, and the need to perform these types of experiments when modeling emotions in different cultural contexts.

When analyzing the results obtained for *Disgust* (Fig. 2.4c), we found that it occupies a large part of the dimensional space. The most likely explanation

for this behavior is that in Spanish, the *Disgust* emotion does not seem to have a clear definition in the dimension of *Arousal* even though it is clear that the *Pleasure* value is negative. This behavior is better observed in Fig. 2.5, where the dispersion of the *Arousal* values for the *Disgust* emotion is very high (but always with a negative *Pleasure* value).

From these results, we estimated the circular normal distribution of each emotion. Fig. 2.6 shows the overlap of the von Mises models obtained for each emotion. As can be observed, there are some emotions that overlap each other. For example, the *Fear* and *Anger* emotions have similar mean angles ($\overline{\alpha}_{fear} = 2.19$ and $\overline{\alpha}_{anger} = 2.42$), and the same relation can be observed for *Boredom* and *Sleepiness*, or *Excitement* and *Happiness*. In addition, as expected, it can be observed that the *Disgust* emotion has a very large standard deviation and is overlapped by several emotions. This representation allows us to estimate the probability of selecting an emotion according to the angle α between the *Pleasure* axis and the line defined by the (*Pleasure*, *Arousal*) point and the origin of the coordinate system. Therefore, instead of defining emotions as areas, we define regions of probability in which emotions can occur with a certain degree of uncertainty. By introducing this uncertainty, the model becomes more versatile, allowing the same point in space to belong to more than one emotion with different probabilities. In addition, with this model, we avoid one of the main problems of a point-based representation because we cannot categorically affirm that a point belongs to one and only one emotion.

We have performed a similar circular statistical analysis of the results obtained by Russell with English speakers [193] (using Formulas 2.1 and 2.2). To demonstrate the need to adapt the emotional models to the language in which it will be used, we have compared these results with those obtained in our experiment with Spanish speakers. In order to determine if the samples are different, we performed a t-Student statistical analysis starting from the null hypothesis H_0: the samples are equal. The results are shown in Table 2.2. As can be observed, the p-value for the emotions *Happiness*, *Excitement*, *Fear*, *Anger*, *Sadness*, and *Sleepiness* is less than 0.05. Therefore, we can reject the null hypothesis H_0 and accept that the samples are significantly different. However, for the *Boredom* and *Calm* emotions, the p-value is greater than 0.05. Therefore, the samples are not significantly different. Fig. 2.7a shows the

(a) Results of our experiment.

(b) Results obtained by Russell.
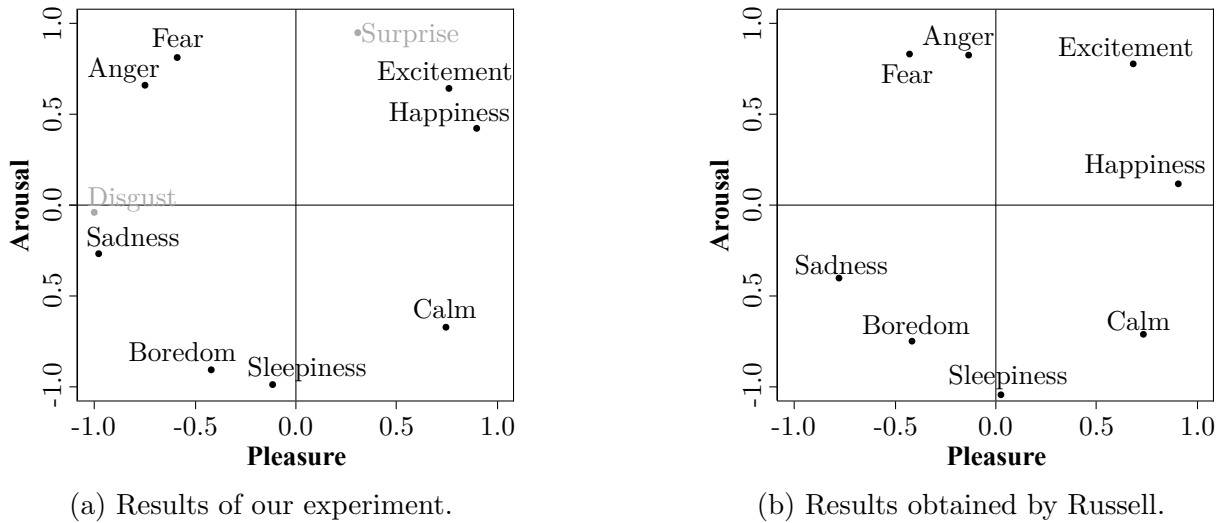
Figure 2.7: Results obtained in our experiment (with emotions expressed in Spanish (see Table 2.1)) and Russell's (with emotions expressed in English) [193]. Note that the emotions not represented in the Russell's experiment are shown in gray.

mean position of the 10 selected emotions in Spanish according to their level of *Pleasure* and *Arousal.* When we compare the results of our second experiment with those obtained by Russell (Fig. 2.7b), some differences can be easily detected. For example, our experiment shows that the *Happiness* emotion is related to higher levels of *Arousal* for European Spanish-speakers than the levels obtained by Russell for British English-speakers. These differences can also be seen in other emotions such as *Anger*, which is more related to lower levels of *Pleasure* in Spanish than in English.

We can conclude that, for six of the eight selected emotions, the differences between languages are significant. Therefore, these results confirm the need for experiments of this type when working with emotions in different cultures and languages since there are significant differences in the affective meaning of the words that are used to define emotions.

Table 2.2: Distribution of emotions in European Spanish and British English.
(Data for British English obtained from [193]). * indicates *p-value* < 0.05.

Emotions in Spanish			Emotions in English			Statistical comparison	
Label/word	Mean $\overline{\alpha}_e$	Sd σ_e	Label/word	Mean $\overline{\alpha}_e$	Sd σ_e	t-value	p-value
Felicidad	0.44	0.33	Happiness	0.14	0.58	3.78	0.0002*
Emoción	0.70	0.18	Excitement	0.84	0.34	-3.27	0.0013*
Miedo	2.19	0.27	Fear	1.27	0.58	12.54	0.0000*
Enfado	2.42	0.30	Anger	1.73	0.76	7.63	0.0000*
Tristeza	3.42	0.40	Sadness	3.62	0.77	-1.98	0.0488*
Aburrimiento	4.28	0.39	Boredom	4.20	0.69	0.91	0.3638
Somnolencia	4.60	0.27	Sleepiness	4.75	0.33	-2.60	0.0102*
Calma	5.55	0.63	Calm	5.52	0.34	0.30	0.7612

2.4 A fuzzy culturally adapted model of affect

We have defined a fuzzy model to represent emotions in the *Pleasure-Arousal*
space using the results of the second experiment. Our model uses a circular
representation in which an emotion is represented by a vector that is defined
by a pair of values of *Pleasure* and *Arousal*. This vector represents both
the type and intensity of the emotion. Taking into account the relationship
between the variables of *Pleasure* and *Arousal* and the intensity of emotions
proposed by Reisenzein [183], we have divided the *Pleasure-Arousal* space into
four fuzzy values of intensity: *Strong*, *Medium*, *Weak*, and *Neutral*. We have
introduced the concept of the *Neutral* intensity to represent those emotions
whose intensity is too low to be considered as elicited.

Therefore, given a point in space defined by its *Pleasure* and *Arousal* value,
we obtain the intensity of the emotion represented by that point using the
modulus of the vector that forms the point with the origin of the coordinate
system. Then, we transform this point into a fuzzy value following the model
of intensity represented in Fig. 2.8, where the direction of the vector indicates
the emotion type. We have used the results obtained in the second experiment
to define the areas where the 10 basic emotions have the greatest probability
of being located. Fig. 2.8 shows the representation of these probability areas

in the circular two-dimensional space. This figure has been obtained from the von Mises model shown in Fig. 2.6 using a single standard deviation. The outer arcs in Fig. 2.8 represent the regions assigned to each emotion. This figure clearly represents the distribution of the basic emotions in the *Pleasure-Arousal* space for Spanish-speakers.

As mentioned in Section 2.3, when we described the model obtained from the results of the second experiment, the distribution of the basic emotions in the *Pleasure-Arousal* space produces overlaps and intersections between emotions. This is easily observed in Fig. 2.8: The *Excitement* emotion intersects with the *Happiness* and *Surprise* emotions. Similarly, the *Boredom* emotion intersects with the *Sleepiness* and *Disgust* emotions. The *Anger*, *Fear*, and *Disgust* emotions have also an intersection. Finally, there is an overlap between the *Sadness* and *Disgust* emotions. This overlap is largely due to the fact that the *Disgust* emotion has a high dispersion. The resulting model clearly shows the relationship between emotions and the angle of the vector that we use to identify the emotion.

Next section presents our model which consists of two processes to represent and express emotions: the *Emotion Representation Process*, a defuzzification process which represents fuzzy emotions in the *Pleasure-Arousal* space; and the *Emotion Expression Process*, a fuzzification process which expresses emotions using fuzzy terms in the same way as human beings express the emotion in their own language.

The Emotion Representation Process

The *Emotion Representation Process* is a defuzzification process, in which the emotions recognized or elicited by an affective agent are internally represented in the *Pleasure-Arousal* two-dimensional space. Human beings express emotions using a fuzzy terminology. Therefore, the model must be able to represent fuzzy emotion values in the *Pleasure-Arousal* continuous space. A fuzzy emotion that is recognized by the agent is internally represented using the mean values of *Pleasure* and *Arousal* obtained in the second experiment. In the same way, the intensity is represented in this multidimensional space by using the mean value of each intensity value from the model represented
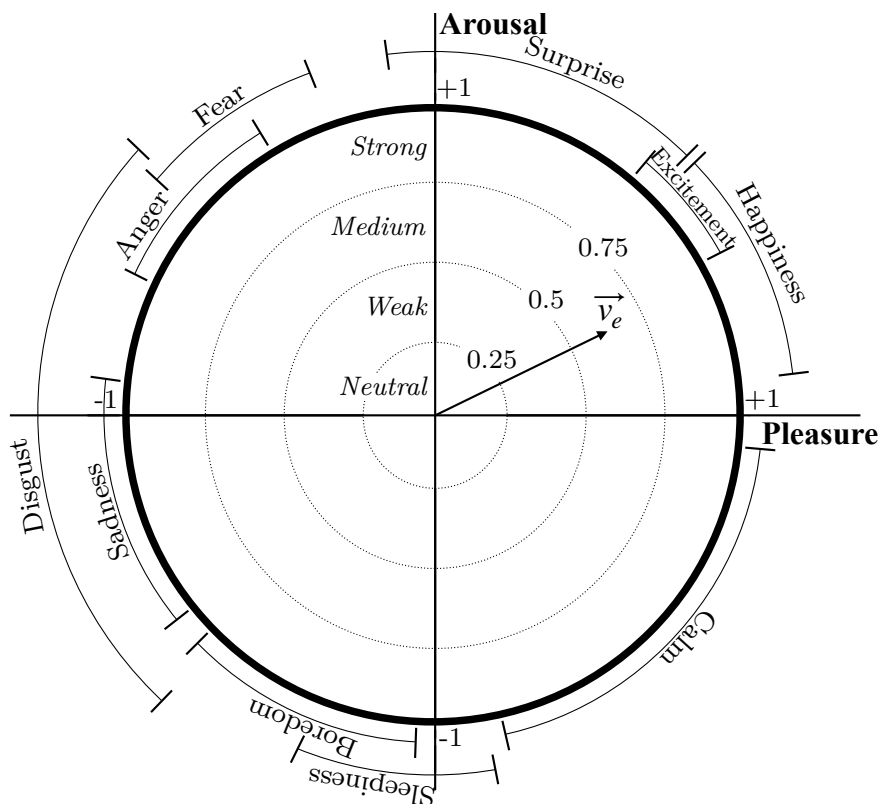
Figure 2.8: Levels of *Pleasure* and *Arousal* associated to Spanish emotions
(see Table 2.1) using one standard deviation.
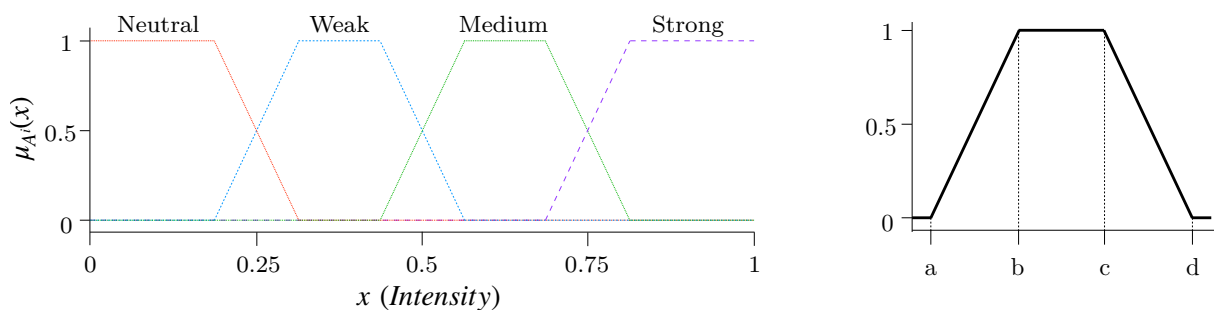
(a) The proposed fuzzy model for the intensity of the emotions
in the *Pleasure-Arousal* space. See Table 2.3 for the linguistic
values in Spanish.

(b) Parameters of the trape-
zoidal membership function.

Figure 2.9: Fuzzy model for the intensity.
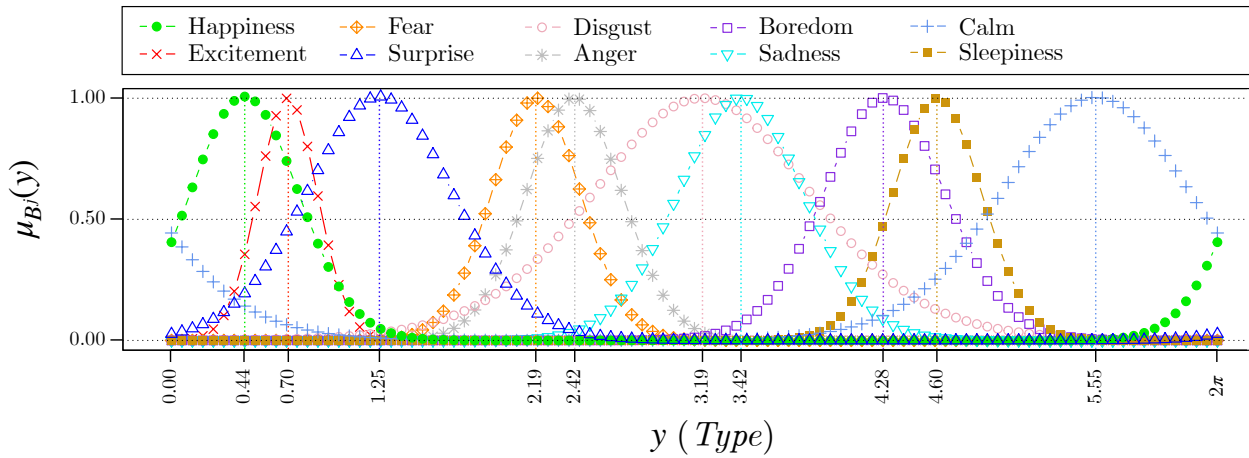
Figure 2.10: Proposed fuzzy model for the type of the Spanish emotions (see Table 2.1) in the *Pleasure-Arousal* space. Note that the order of the means of the emotions on the plot corresponds to the order of the legend.

in Fig. 2.8. Let us illustrate this process with a simple example. If the emotion perceived by the emotion recognition process of an affective agent is *Happiness* with a *Medium* intensity, the agent will internally represent this emotion using a vector $\vec{v_e}$ with a direction $\alpha_e = 0.44$ radians (≈ 25.2 degrees) (see Table 2.1) and a modulus $|\vec{v_e}| = 0.625$, which correspond to the *Medium* intensity (the midpoint between 0.5 and 0.75 according to Fig. 2.8). From this vector $\vec{v_e}$ representing the emotion e, the agent will obtain the *Pleasure* P_e and *Arousal* A_e values using the following equation:

$$(2.3) \qquad (P_e, A_e) = (|\vec{v_e}| \cdot \cos \alpha_e, |\vec{v_e}| \cdot \sin \alpha_e)$$

Therefore, by using this formula, the values of *Pleasure* and *Arousal* associated with the "*Happiness* emotion with *Medium* intensity" in the Spanish language are 0.57 and 0.27, respectively. An example of the resulting representation of the vector $\vec{v_e}$ in our model is shown in Fig. 2.8.

Note that this representation of the emotion is based on the values of *Pleasure* and *Arousal*. Therefore, this representation has a universal semantics [107]. The agent uses this representation of the emotion for all its affective processes. When the agent needs to express an emotion, which is internally

Table 2.3: Parameter ranges for the intensity membership function.

Intensity level	Parameters			
in Spanish (English)	a	b	c	d
Neutral (Neutral)	0	0	0.2	0.3
Bajo (Weak)	0.2	0.3	0.45	0.55
Medio (Medium)	0.45	0.55	0.7	0.8
Alto (Strong)	0.7	0.8	1	1

Table 2.4: $\mu_{A^i}(u)$ values for the example ($u = 0.625$).

	Neutral	Weak	Medium	Strong
$\mu_{A^i}(0.625)$	0	0	1	0

Table 2.5: $\mu_{B^j}(t)$ values for the example ($t = 0.44$). For the Spanish labels see Table 2.1.

	Happiness	Surprise	Fear	Anger	Disgust	Sadness	Boredom	Sleepiness	Calm
$\mu_{B^j}(0.44)$	1	0.35	0.21	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0

represented in the *Pleasure-Arousal* space, the agent will use the process presented in the next section.

The Emotion Expression Process

Any emotion that is internally represented by a vector $\vec{v}_e$ is transformed into a fuzzy expression (emotion label and fuzzy intensity) by the *Emotion Expression Process*. This process allows the agent to express its emotions using a fuzzy expression adapted to the cultural environment in which the agent is located. In this process, a fuzzification of the vector $\vec{v}_e$ is performed transforming the intensity and the type of the emotion into fuzzy values. This process is composed of two fuzzification subprocesses: One for the intensity and the other for the emotion type.

We define the linguistic variable x for the fuzzification process of the in-

tensity of the emotion as a quintuplet [239]:

$$(2.4) \qquad x = \langle L_x, T(L_x), U_x, G_x, M_x \rangle$$

where L_x represents the name of the linguistic variable $L_x = intensity$, and $T(L_x)$ represents the set of fuzzy terms (linguistic values for the intensity) that the variable x can take:

$$(2.5) \qquad T(intensity) = \{Strong, Medium, Weak, Neutral\}$$

U_x (universe of discourse) represents the range of crisp values that the variable can take in the range $U_x = [0, 1]$, G_x is the syntactic rule that generates the terms in $T(intensity)$, and M_x is a semantic rule that associate each linguistic term in $T(intensity)$ with its meaning.

We define a Type-1 fuzzy logic set $\tilde{A} \in M_x$ using a trapezoidal membership function defined by the equation [126, 11]:

$$(2.6) \qquad \mu_{A^i}(u) = max\left(min\left(\frac{u-a}{b-a}, 1, \frac{d-u}{d-c}\right), 0\right)$$

where $i \in T(intensity)$, A^i represents the ith fuzzy set belonging to $\tilde{A}$, u takes the value of the modulus of the vector of the emotion such that $u \in U_x$, and a, b, c, and d are the parameters of the membership function (see Fig. 2.9b) that is described in Table 2.3 [11]. Fig. 2.9a shows the resulting fuzzy model.

On the other hand, we define the linguistic variable y for the fuzzification process of the emotion type ($L_y = type$):

$$(2.7) \qquad \begin{aligned} T(type) = \{&Happiness, Excitement, Surprise, Fear, Anger, \\ &Disgust, Sadness, Boredom, Sleepiness, Calm\} \end{aligned}$$

The universe of discourse is defined in the range $U_y = [0, 2\pi]$. We use the results obtained in the second experiment (Fig. 2.6) to define a Type-1 fuzzy logic set $\tilde{B} \in M_y$. The proposed membership function for an emotion t is derived from the normalization of the von Mises distribution and is defined as:

$$(2.8) \qquad \mu_{B^j}(t) = \frac{m_j(t) - \min_{x \in [0,2\pi]} m_j(x)}{\max_{x \in [0,2\pi]} m_j(x) - \min_{x \in [0,2\pi]} m_j(x)}$$

where $j \in T(type)$, B^j represents the jth fuzzy set belonging to $\tilde{B}$, t takes
the value of the angle of the vector of the emotion expressed in radians such
that $t \in U_y$, and $m_j(\theta)$ represents the jth von Mises function defined as:

$$(2.9) \qquad m_j(\theta) = \frac{\exp\left(\kappa_j \cos(\theta - \overline{\alpha_j})\right)}{2\pi I_0(\kappa_j)}$$

where $\overline{\alpha_j}$ is the mean for the jth fuzzy set B^j (see Table 2.1), κ_j represents
the jth concentration parameter of the von Mises distribution estimated as
the inverse of the variance $(1/\sigma_j)$ [102], and $I_0(\kappa_j)$ is the modified Bessel
function of order zero used to normalize the function [92]. Fig. 2.10 shows
the resulting fuzzy model.

As a result of this process, an affective agent can express an emotion,
which is internally represented in the *Pleasure-Arousal* space by a vector $\vec{v}_e$,
by using the same fuzzy expression (emotion label and intensity) that humans
use in their language.

Let us again consider our example of an affective agent that has an emo-
tion represented internally by the vector $\vec{v}_e$ (Fig. 2.8). The *Pleasure* and
Arousal coordinates of vector $\vec{v}_e$ are 0.57 and 0.27, respectively. Therefore,
the direction of the vector is 0.44 radians and the modulus is 0.625. When
the affective agent needs to express this emotion in a specific language, it uses
the fuzzification process describe above. This fuzzification process for the in-
tensity produces the results shown in Table 2.4. Therefore, the selected fuzzy
value for the intensity is *Medium*. On the other hand, the fuzzification process
for the emotion type produces the results of Table 2.5. If the agent wants
to express only one emotion, the agent will select the emotion term/label
with the maximum μ_{B^j}, which in our example is *Happiness*. Therefore, the
Emotion Expression Process will generate the fuzzy emotion *Happiness* with
Medium intensity as a result of the fuzzification of the emotion represented
in the multidimensional space by the vector $\vec{v}_e$.

This model can be easily adapted to other languages and cultures. Our
model allows a multicultural affective agent to represent and express an emo-
tion using the same interpretation of the fuzzy emotional terms used in the
language and culture of its human interlocutor. This can be achieved by
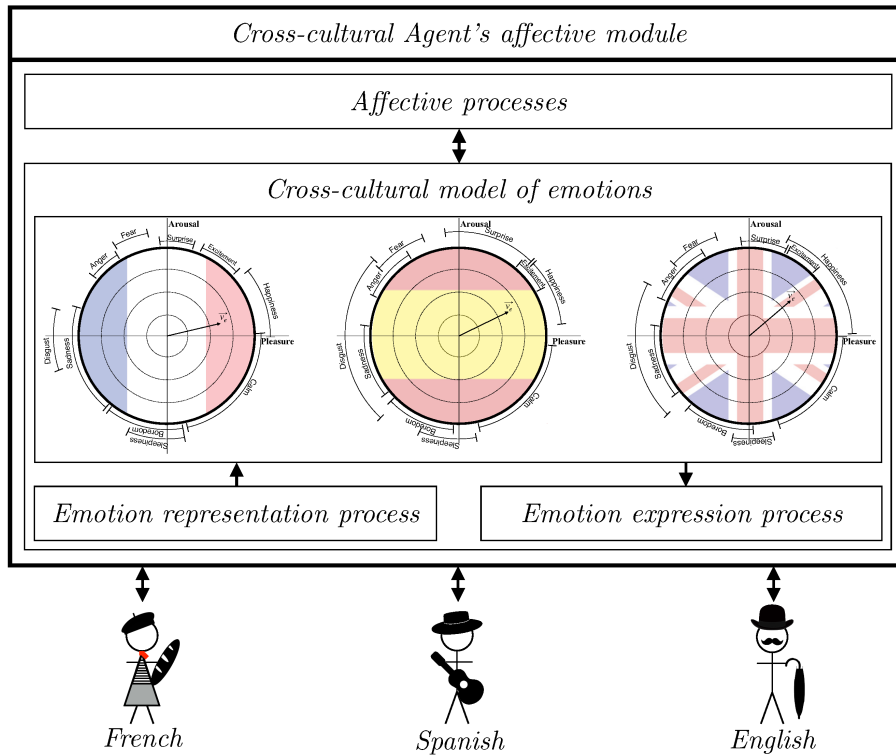
Figure 2.11: Example of a multicultural affective agent module.

defining a specific dimensional *Pleasure-Arousal* space for each cultural environment in which the agent is located. All these two-dimensional spaces can be easily incorporated to the affective component of the multicultural affective agent (Fig. 2.11).

2.5 Conclusion and future work

In this paper, we have proposed an experiment-based method to adapt the emotion representation to the culture and the language in which the emotion will be used. The results of the first experiment have corroborated the importance of adapting the translation of the emotional models taking into account the different meanings that humans give to emotion labels in different cultures. As observed in the translation of the *Misery* emotion, the literal translation does not have the same contextual meaning for European

Spanish-speakers and British English-speakers. The results of the second experiment have allowed us to assign different areas in the *Pleasure-Arousal* space to the different basic emotions. When comparing the results obtained in our experiment with those obtained for British English-speakers, it can be observed that the emotion labels have different values for *Pleasure* and *Arousal* depending on the language. This confirms the need for performing this adaptation with experiments when emotional models are used in other cultures and languages.

From the results obtained in the experiments, we have defined a fuzzy model that represents emotions using a *Pleasure-Arousal* continuous space instead of simple labels. With this cross-cultural dimensional representation, an emotion can be defined as a vector in which the modulus represents the intensity of the emotion, while the type of emotion depends on the direction of this vector. Our model improves the typical representation of emotions based on labels because it allows emotions and their intensity to be easily represented and modified using a dimensional representation.

We have also defined a fuzzification process that determines an emotion label in a specific language from a pair of *Pleasure* and *Arousal* values. Thus, using our model, an agent can express emotions using the same fuzzy terminology as humans.

We are currently incorporating this model into the affective agent architecture $GenIA^3$ to improve the simulation of the different affective processes. In addition, we want to analyze whether this emotion representation can be improved by introducing a third dimension (*Dominance*) to better identify emotions such as *Fear* or *Anger* and to analyze the influence of this dimension on social behavior [124].

Computational affective knowledge representation for agents located in a multicultural environment

Abstract

Emotions are present in all human interactions. However, there are differences between cultures in how emotions are expressed in interaction. Consequently, when interactions are translated into another language, this culturally relevant information can be lost. In this paper, we show the results of two experiments conducted to adapt a new model of affect representation, based on a polar coordinate system in a space composed of pleasure and arousal dimensions, to the Portuguese and Swedish languages. In the first experiment, we determine the most appropriate labels to define the pleasure and arousal dimensions in each language. In the second experiment, we identify the most suitable values of pleasure and arousal dimensions for each emotion in Portuguese and Swedish. The results obtained in these experiments are compared with a previous experiment conducted with European Spanish-speaking participants. The analysis of the results shows the importance of adapting the affect representation models to the cultures and languages in which they will be used. We demonstrate significant differences in the values of pleasure and arousal associated with emotions across languages and cultures. The results also show that there are not significant differences in gender or age when associating levels of pleasure and arousal to emotions.

3.1 Introduction

When analyzing human behavior, it is common to find situations in which emotions prevail over logic, conditioning our capacity for judgment and decision making and altering our behavior. Therefore, considering affective states in software processes can improve the simulation of human behavior, the simulation of human social organizations, and human-computer interaction. In fact, systems that are capable of recognizing and simulating affective states have proved to be more credible and more reliable, thus reducing the frustration and stress of human users and improving human-machine interaction [165].

Human affective behavior encompasses a set of constructs and cognitive processes that make it complex to simulate using software agents. One of the main constraints on simulations comes from the way in which humans interpret affective knowledge. To express, represent, and interpret the affective knowledge, humans typically use different abstractions and metaphors. For example, through the use of words or labels such as "Sadness" or "Joy". This way of representation is commonly used in the area of affective computing [170] to generate models for recognition, interpretation, processing, and simulation of affective states [17, 20, 74]. However, the use of words or labels is a simplification that does not allow to represent all the information inherent to affective states such as the intensity with which an affective state is perceived. An alternative to this representation is the use of models based on different cognitive dimensions. One of the most common dimensional models in literature is the *Core Affect* model [194]. That model defines affective states through two dimensions: pleasure and arousal. Dimensional models have been widely used in affective computing [6, 81]. Still, existing proposals also simplify affective states, losing some of the information inherent to affective states. Moreover, it should be noted that, being cognitive constructs associated with different words, culture and language condition the representation, expression, and interpretation of affective states. Representing affective states while capturing cross-cultural differences poses a challenge since, the cognitive conceptualization of an affective state, may vary across cultures and languages thus influencing the levels of pleasure and arousal as-

sociated with that affective state. The use of representation models that are
not adapted to the culture and language may produce misunderstandings and
errors that compromise the simulations of affective behavior and users' expe-
rience. Therefore, establishing representation models capable of adapting to
culture and language is a step towards in improving the simulation of human
behavior and the personalization of human-machine interaction [23, 196].

In this paper, we present the results of two experiments conducted with
Portuguese and Swedish participants to adapt a pleasure-arousal representa-
tion space to both cultural environments. Based on the results obtained, we
defined the probability models for each affective state label. These probability
models can be used to represent recognized affective states in the pleasure-
arousal space and allow software agents to express their affective states in the
same terms used by humans in different cross-cultural environments. Then,
we compare the results with the previous study conducted with European
Spanish speakers. We finally present a computational model of affect adapted
to Portuguese and Swedish and a potential application of this cross-cultural
model in the area of human-aware planning.

The rest of this paper is organized as follows. In Section 3.2 we present the
most relevant theories and models for the representation of affective states.
Section 3.3 presents our model of affect representation. In Section 3.4 we de-
scribe the methodology used in the experiments conducted to determine the
levels of pleasure and arousal associated with affective states in Portuguese
and Swedish. The results obtained in the experiments are shown in Sec-
tion 3.5. Then, a discussion of the results obtained along with the definition
of the cross-cultural model of affect is provided in Section 3.6. Section 3.7
proposes potential applications for our representation model of affect. Finally,
the main conclusions and some future works are presented in Section 3.8.

3.2   Related work

Affective states are constructs that define different experiences, including
emotions and moods. On the one hand, emotions can be understood, in
general terms, as immediate responses to a specific stimulus [140]. On the

other hand, moods are described in the literature as affective responses that have a longer duration than emotions, ranging from minutes to days or weeks [90]. In addition, mood is not necessarily related to a specific stimulus [44]. Instead, moods are generally triggered by the achievement of a set of cognitive and affective processes.

From the area of affective computing [170] different approaches focused on the recognition, interpretation, processing, and/or simulation of different emotional behaviors and affective states have been conducted [35, 6, 221]. However, not much attention has been paid to the development of models to represent the knowledge related to affective states specifically designed for its use in computational systems. To represent and express the complexity of knowledge related to affective states, humans usually resort to the use of different metaphors and abstractions. One of the most recurrent classifications in the literature divides these metaphors and abstractions into two approaches: categorical and dimensional [93].

Categorical approaches are based on the use of a limited set of words or labels to define the different affective states [66]. These labels provide a means to verbalize affective states, which is a key factor in human affective interactions. One of the best known models is the model of *Basic Emotions* proposed by P. Ekman in which six labels are defined: "Sadness", "Happiness", "Anger", "Fear", "Disgust", and "Surprise" [65]. This model is one of the most widely used in the area of affective computing to express, recognize, interpret, and represent affective states [20, 173, 74]. For example, in [17], a model of emotion recognition is presented. The authors propose to use the six basic emotions proposed by Ekman to classify emotions recognized in text. However, by a label based representation we assume a high level of abstraction since we condense the entire spectrum of affective states into a limited set of labels. Therefore, we are losing part of the affective information, such as the intensity of the affective state, or the probability of moving from one affective state to another depending on a certain stimulus [71, 163, 232]. For example, if a person is "Sad" and receives good news, the probability of transitioning to a positive affective state such as"Joy" may be lower than if the person is "Happy". In addition, this representation assumes that labels/words representing emotions have the same meaning in all cultures and languages.

Nonetheless, depending on the model of emotional representation being used,
translating emotions from one language to another can lead to cultural mis-
understandings since emotions are not always understood in the same way
in different cultures and languages [196]. In fact, there are some labels that
express affective states that are very difficult to translate from one language
to another. For example, the German vocabulary has an extensive catalog of
affective labels (or words) defining emotions such as "Schadenfreude", which
can be summarized as "pleasure from the suffering of others", or "Sehnsucht",
which can be described as "a strong desire for an alternative life", but whose
overall conceptualization is very difficult to describe by means of a word or
label in other languages such as English or Spanish [107, 137].

Theories based on dimensional approaches allow reducing the level of ab-
straction by representing affective states through different dimensions. These
theories have their origin in psychological constructivism, according to which
a finite set of labels cannot establish affective states since they depend on the
individual's subjective experience that occurs in a particular context (e.g., a
particular culture or language) [151, 101]. One of the most representative di-
mensional models is the *Core Affect* model proposed by J.A. Russell [194]. In
that model, affective experiences are defined as emotional cues and are repre-
sented by two dimensions: pleasure and arousal. More specifically, emotions
can be defined as high or low arousal (relative to the emotional arousal dimen-
sion) and positive or negative (considering the emotional pleasure dimension).
Russell also evidenced through experimentation that the labels used to de-
fine the different affective states followed a circular pattern along these two
dimensions. From these results he developed the *Circumplex Model of Af-
fect* [174]. The *Circumplex Model of Affect* uses the multidimensional scaling
of labels or words that define emotions [193], concluding that the emotional
pleasure dimension and the emotional arousal dimension are the ones that
best support the judgments people make about emotions [192, 68]. In that
model, all the labels referring to the different affective states are represented
as points composed of pleasure and arousal components and are organized
in a circular way around these two dimensions, conceptualizing emotions as
being systematically related to each other [223], which is also the idea de-
fended by H. Scholesberg [202]. Being a representation based on continuous

dimensions, it is possible to establish relationships between the distribution of the different affective states. For example, it can be assumed that, in general, affective states related to the word "Joy" are related to positive values of the pleasure and arousal dimensions, while those related to the word "Sadness" are related to negative values on these two dimensions.

Some authors have discussed the number of dimensions that define affect. For example, in [154], it is proposed to introduce the *dominance* dimension to the pleasure-arousal space resulting in a model known as the PAD model. Dominance dimension allows to disambiguate some emotions that are overlapped when only the dimensions of pleasure and arousal are used. Nevertheless, some studies show that this dimension does not have a substantial impact on the representation of affect across cultures [107] and that humans exhibit difficulties and confusion when they try to assign a dominance value to the emotions that they are feeling [122].

In addition, dimensional models, such as the *Circumplex Model of Affect*, make it possible to highlight the differences that appear across cultures. Over the years, different authors have shown, through experimentation, the existence of variations in the levels of pleasure and arousal associated with each affective state in different cultures and languages [196, 23, 46]. For example, the study presented in [107], analyzed the discrepancy of 24 emotional labels in about 2500 languages. That study concludes that there are significant variations between the values of pleasure and arousal across cultures. Therefore, the direct translations of emotion labels can be subject to bias because the meaning of the words can differ between languages. The authors also evidenced that the differences become smaller the closer the geographical proximity of the language families is.

The circumplex model of affect has also been used in affective computing [243, 204]. For example, in [136], a model for estimating the facial expressions of a humanoid robot is presented. Authors use the *Circumplex Model of Affect* to represent emotions as points composed by the pleasure and arousal dimensions. However, the use of the *Circumplex Model of Affect* in affective computing also presents some disadvantages. Let us suppose that we define affective states labels by a single point in the dimensional representation space, and then we generate a random point in that space. If the point does

not match with any of the points predefined for each affective state label, it
is not possible to determine the label associated with that point. A possible
approach, used in some proposals [118, 131], consist in estimating the dis-
tance of a random point from each of the predefined points for the affective
labels in the model and then, assign the nearest affective label. By this ap-
proach it is necessary to assume that the distribution of affective labels in
the pleasure-arousal space is equal and uniform. Although, this way of rep-
resenting affective labels is artificial and is far from how affective states are
distributed throughout the pleasure-arousal space as demonstrated in work
presented at [220]. In addition, in a point-based approach, it is complex to
assign a level of uncertainty that a random point belongs to a certain label
or represent the transitions between two affective states. It is also difficult
to simulate other affective behaviors such as the influence on the affective
state of other affective factors like personality [216]. Still, in a point-based
representation, the intensity of affective states is not considered [34]. Despite,
there is empirical evidence that establishes a direct relationship between the
levels of pleasure and arousal and the intensity with which different affective
states are experienced. This means that, values close to zero in both dimen-
sions would indicate very low or no intensity. In contrast, values close to one
would indicate high intensity [183].

A possible alternative to the point-based representation is the vector-based
representation. In this representation, the direction of the vector is used to
determine the label that represents a point in the pleasure-arousal space. For
example, in [111], a vector-based representation using pleasure-arousal space
to express emotions in a robot is proposed. In that model, each emotion is
represented by a uniform range of fifteen degrees. Note that, the emotion
"Glad" is defined in the range from 0 to 14 degrees while "Happy" is defined
in the range from 15 to 29 degrees. Therefore, that representation presents the
same problem as the point-based representation, since the way in which the
variations in the dispersion of affective states are represented is symmetrical
and, consequently, far from the mental representation that humans have of
affective states.

Finally, there seems to be a trend towards the development of affective
models based on fuzzy logic [67, 116, 108]. The fuzzy logic allows adding

uncertainty to the representation model by allowing a point in the pleasure-arousal space to belong to more than one label. Moreover, fuzzy logic allows us to express affective states in the same terms that humans use (e.g., "very sad" or a "little happy"). For example, in [121], a fuzzy logic-based model for translating emotional labels from Spanish to English is proposed. A methodology based on an experiment in which 8 Spanish speakers assigned values of pleasure, arousal and dominance to 30 Spanish emotional labels/words was proposed to establish the discriminant values of the fuzzy membership functions for each emotion. The results were compared with the models obtained in the previous experiment conducted with English-speaking participants presented in [123]. Through this comparison, a set of fuzzy functions to translate the emotional labels between the Spanish and English languages were obtained.

3.3 A cross-cultural model of emotions for affective computing

In [220], we present a new cross-cultural model of emotion representation for affective computing. Our model differs from the representations discussed in Section 3.2, by providing a new system for representing affective states based on the use of probability areas to define each affective state in the pleasure-arousal space. By using probability areas we add a certain level of uncertainty, since each point in the space composed of the pleasure and arousal components has a certain probability of belonging to each affective label. Our model follows a vector-based representation in a polar coordinate system in which the direction indicates the probability of experiencing a particular affective state and the modulus indicates its intensity. However, in a polar coordinate system, traditional linear statistical models may not be suitable since they do not satisfy the linearity constraint because they are constrained between 0 and 2π. The main problem underlying the application of linear statistics arises when studying the probability regions near 0 and 2π since, in a linear model, these two regions would be independent but, in a polar coordinate system, these two regions are dependent. An approach

to address the problem of linearity can be found in [106], where a model to represent emotions using a vector-based representation in the pleasure and arousal dimensions is proposed. In that proposal, the representation space is modified by displacing the coordinates' origin to the negative extreme of the arousal dimension. As a result, the model is bounded between 0 and π, thus reducing the problem that appeared in a representation based on a complete polar coordinate system. Still, this representation has two disadvantages: on the one hand, there is a decontextualization of the sense of negative arousal, which is related to emotions such as "Sleepiness", "Sadness", "Boredom", or "Calm". On the other hand, this representation modifies both the modulus of the vector and its direction, since the origin of coordinates of the vector is located in the negative component of the arousal dimension. By modifying the modulus of the vector, the relationship between the intensity of the affective states and the values of the pleasure and arousal components is blurred. To avoid these issues we propose to replace the common linear statistics with circular statistics [18]. Circular statistics are designed for data that present a circular distribution, as is the case of affective states across the dimensions of pleasure and arousal. The most suitable approach to estimate the dispersion in a circular representation space composed of the pleasure and arousal dimensions is through the circular standard deviation, which is defined as follows [169, 109]:

$$(3.1) \qquad\qquad \sigma = \sqrt{-2 \, \ln \overline{R}}$$

where

$$(3.2) \qquad\qquad \overline{R} = \sqrt{\overline{P}^2 + \overline{A}^2}$$

where $\overline{P}$ and $\overline{A}$ represent the arithmetic mean for the pleasure and arousal dimensions respectively.

On the other hand, to estimate the probability distribution in a circular representation space the von Mises distribution [102] can be considered. This
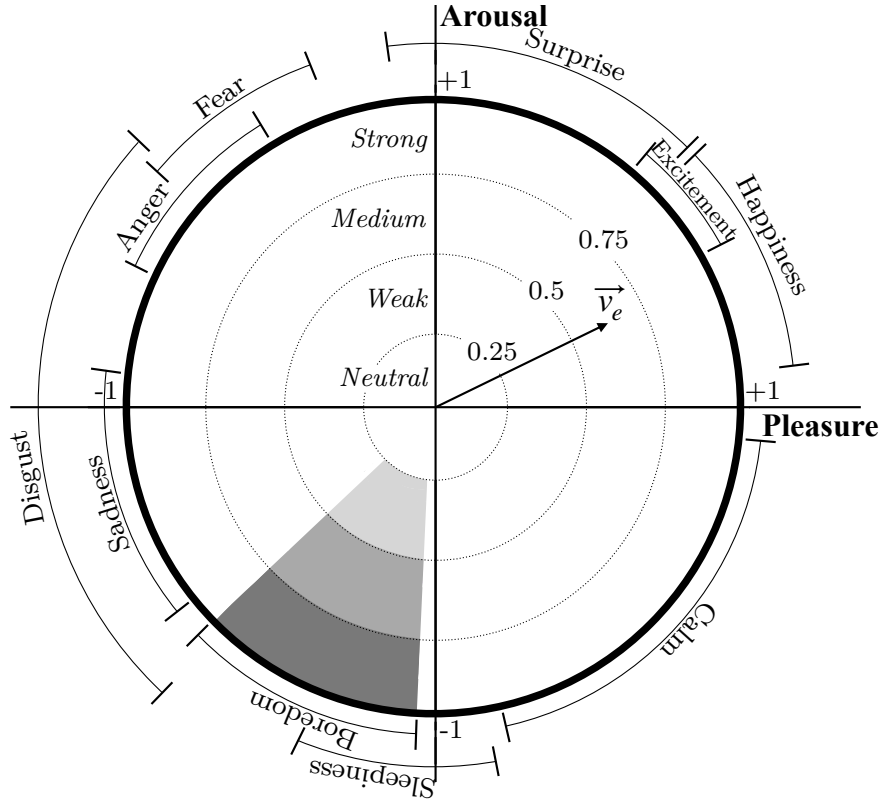
Figure 3.1: Distribution of probability functions in the pleasure-arousal space using one standard deviation for Spanish.

distribution allows estimating the Gaussian (or normal) distribution in circular data. Given an angle α, the von Mises probability density function is defined as follows:

$$(3.3) \qquad f(\alpha \mid \mu, \kappa) = \frac{e^{\kappa \cos(\alpha - \mu)}}{2\pi I_0(\kappa))}$$

where μ represents the mean of the distribution, κ represents the concentration parameter of the distribution which is calculated as the inverse of the circular variance $(1/\sigma^2)$, and $I_0(\kappa)$ is the modified Bessel function of order zero used to normalize the function [92].

Figure 3.1 shows an example of our model of emotion representation adapted to European Spanish-speakers. In this example, the outer rings show the probability areas of each of the ten emotions using one circular standard deviation,

the vector $\vec{V_e}$ represents an affective state, and the inner rings represent four regions of intensity: *Strong, Medium, Weak,* and *Neutral.* The grey shaded area shows an example of the probability region related to the emotion "Boredom" using one circular standard deviation. Note that the different grey tones represent the different intensity values for the "Boredom" emotion. We can represent the different affective states through this model, abstracting from the labels that represent each affective state, which will only be necessary when a software agent needs to express, interpret, or represent an affective state. Vectors can be modified over time by the occurrence of a sequence of events that happen to an agent. This makes it possible to evaluate the relationship between previous and current affective states, allowing the simulation of different human affective behaviors. For example, if the agent's has an affective state at instant t_n and receives a new positive event at instant t_{n+1}, we can adapt the agent's emotional response to the affective state it had in t_n by modifying the intensity of the emotion generated by the positive event in t_{n+1}.

Although the representation of affective states through the dimensions of pleasure and arousal can be considered universal [174, 107], probability areas associated with each affective state depend on the culture and language in which the affective state is needed to be interpreted, expressed, or represented. The direct use of a dimensional model without adapting it to the cultural environment in which it will be used, can produce misunderstandings in the process of selection or representation of the emotion, resulting in erroneous behaviors that will reduce the realism of the simulation of the emotional behavior. Therefore, in our design, we proposed an experiment-based methodology to establish the levels of pleasure and arousal for each emotion in the language in which the model is to be used. Our experiment-based methodology can easily modify these probability areas to include different emotions and adapt the model to the cultural environment in which it is being used.

In addition, through this multidimensional cross-cultural model of affect, an agent can have more than one probability model for different languages and adapt its model of representation and expression of affective states to the language and culture of its human interlocutor. To do that, we establish

two processes based on a fuzzy logic approach: the *Emotion Representation Process*, which is a defuzzification process, which allows representing words related to affective states in the pleasure-arousal space adapted to the cultural environment; and the *Emotion Expression Process*, that is a fuzzification process, which determines an emotion label in a specific language from a pair of pleasure and arousal values [220]. These processes allow the agent to express its affective state and represent the perceived affective state using the same interpretation of the fuzzy emotional terms used by its human interlocutor and adapted to his/her cultural environment.

3.4   Experiment design

To find the areas associated with each affective state word or label in a target language, we used a methodology based on two experiments: in the first experiment, the words used to represent dimensions of pleasure and arousal are translated into the target language; then, the results of the first experiment are used in the second experiment to obtain the levels of pleasure and arousal associated with each affective state in the target language. In this paper, we present the results of applying this methodology to define computational models for representing affective states in Portuguese and Swedish.

Methodology of the first experiment

The current study aims to minimize bias in the labelling of emotions, while recognizing prior research results. In the literature there are different references to the pleasure-arousal dimensions. Generally, these dimensions are defined by four affective categories: *pleasure* and its opposite *misery*, and *arousal* and its opposite "Sleepiness" [193]. However, as stated in Section 3.2, the terms used to define affective states or categories are culturally and linguistically dependent. Therefore, using a literal translation of these terms may result in cultural misunderstandings, as the term associated with the translation may not fully reflect the original meaning of an affective category label. This first experiment is designed to identify the labels that best repre-

sent the meaning of the pleasure and arousal dimensions in a target language
(in this work the target languages are Portuguese and Swedish).

The problem of translating into other languages the terms that humans
associate with affective states and affective categories has been explored for
years in the literature [23, 46, 107]. For example, a prior study presented in
[132], observed interpretational issues concerning the naming of basic dimen-
sions in the previous pleasure and arousal models, and suggests a labelling
system for the pleasure-arousal space consisting of forty-eight emotion adjec-
tives in English. A proceeding study by I. Knez and S. Hygge [127] evaluated
a Swedish translation of the forty-eight emotion adjectives observed a rea-
sonable consistency with earlier pleasure and arousal models. Similarly, the
studies conducted in [48, 143] proposed some adjectives to translate these
affective categories in specific domains. These adjectives can thus be a good
starting point for the present study.

Hypothesis

The main hypothesis to be validated by this first experiment is that the
literal translation of the labels associated with the dimensions of pleasure
and arousal may not be adequate across different cultures and languages.
The use of a single translated label may not encompass the original global
construct.

Participants

Participants were recruited through Prolific [166]. Prolific is an online plat-
form for conducting experiments that handles the recruitment of participants
and the acquisition of participants' personal information such as their nation-
ality, age, and gender. The platform provides different filters to establish the
best suited participants for each experiment. When participants successfully
complete the experiment Prolific handles the payment. In case participants
do not complete the experiment properly (e.g., they fail attention check ques-
tions) they suffer a penalty that affects their eligibility for other experiments
and they do not receive any economic reward.

In this first experiment, we required participants' nationality, place of residence, and place of birth to be either Portuguese and Swedish in each case. In addition, we also specified that the first language should be Portuguese or Swedish in each case and that the age should be over 18 years old. Finally, we set an approval rate of over 90%. This approval rate measures the percentage of studies successfully completed by the participants and, in Prolific, is used as a measure of the quality of the participants.

For this first experiment, 50 participants were recruited for each of the target languages. In the Portuguese experiment, the sample was divided into 15 women and 35 men ranging in age from 18 to 67 years old ($\mu = 26.28$, $\sigma = 9.98$). The mean of the approval rate was situated at 99.36 with a minimum of 95 and a mean of 73.84 studies performed successfully.

For the Swedish experiment, the sample was composed of 13 women, 35 men, and 2 participants who preferred to not indicate their gender. In this case, the age of the participants ranged from 18 to 65 years old ($\mu = 29.96$, $\sigma = 9.99$). The mean of the approval rate for Swedish participants was situated at 99.50 with a minimum of 95 and a mean of 115.32 studies performed successfully.

Materials

To translate the four affective categories associated with the dimensions of pleasure and arousal (i.e., *pleasure* and its opposite *misery*, and *arousal* and its opposite "Sleepiness") into the target languages (i.e., Portuguese and Swedish), we followed the methodology presented in [220]. According to this methodology, for the first experiment (concerning identification of suitable pleasure-arousal end-points), a set of labels was chosen for each language. This labels were selected from the common Portuguese and Swedish labels proposed in [48, 143] and [127] respectively. With the aim to minimize bias, we complemented both set of labels with different synonyms collected from the dictionaries: Priberam [177] and Infopedia [105] for Portuguese and Bab.la [12] and Synonymer [214] for Swedish. Synonyms were included as found in the dictionaries, while excluding synonyms that did not fit with the arousal and pleasure end-points. Finally, from these research, we select 28 words

for Portuguese and 30 words for Swedish. Then we divided these words into four sets of words, approximately 5-9 words in each set for Portuguese (see Figure 3.2) and with approximately 7-9 words in each set for Swedish (see Figure 3.3), representing end-points (low/high pleasure and low/high arousal).

To explore the words that participants related most closely to each of the four affective categories in each target language, we designed an online questionnaire that was composed of four sections, one for each affective category. In each section, two checkbox questions were designed to identify the best words to translate each affective category. The first question asked each participant to read a set of words and select the word (or words) that clearly did not express the same meaning as the rest of the words in the set. In the second question, the same set of words was shown and each participant was asked to read the set of words and select the word (or words) that best represented the concept of that set.

We added an extra section containing an attention check question. Following the same structure as the questions in the previous sections, a set of six words was shown, five of which represented color names and one a name of an animal. Each participant was asked to carefully read the set of words and to select the word (or words) that did not represent the same concept as the rest of the set.

Procedure

In order to register for the questionnaire, participants had to enter their Prolific ID. At the beginning of the experiment, participants were provided with the instructions describing the task to be performed along with two examples for illustrative purposes. Then, participants were asked to complete the four sections of the questionnaire without a time limit.

Methodology of the second experiment

In the second experiment (concerning definition of emotions in terms of pleasure-arousal), the aim was to detect the regions of the pleasure-arousal space associated to each emotion label in Portuguese and Swedish. Initially, for this

experiment we selected the ten emotions proposed in [196]: "Fear", "Surprise", "Disgust", "Anger", "Boredom", "Sleepiness", "Sadness", "Excitement", "Calm", and "Happiness". Similarly to the first experiment, suitable labels for representing emotions were chosen that are common with the Swedish labels of Knez and Hygge [127] and with the Portuguese labels presented in [48] and [143], coupled with the resulting end-point labels from the first experiment. We decided to select for Portuguese the words: "Medo", "Surpresa", "Aversão", "Raiva", "Aborrecimento", "Sonolência", "Tristeza", "Excitação", "Calmo", and "Felicidade". On the other hand, for Swedish, we decided to select the words: "Rädsla", "Överraskad", "Avsky", "Ilska", "Uttråkad", "Trött", "Ledsen", "Spänning", "Lugn", and "Lycka". Due to our approach to minimize bias while recognizing prior research, we believe a suitable set of labels is selected.

Hypothesis

The main hypothesis for this second experiment is that the levels of pleasure and arousal associated with an emotional label depends on the culture and language. Therefore, the same pleasure and arousal values cannot be used to represent emotion words/labels translated from one language to another.

Participants

In this second experiment, the Prolific platform was also used to recruit participants by applying the same demographic, language, age, and approval rate filters as in the first experiment.

In the experiment with Portuguese participants, 50 women and 50 men ranging in age from 18 to 52 years old ($\mu = 24.10$, $\sigma = 5.91$) were recruited. The mean of the approval rate for Portuguese participants was 99.06 with a minimum of 95 and a mean of 82.44 studies performed successfully.

For the Swedish experiment, 50 women and 50 men ranging in age from 18 to 67 years old ($\mu = 30.79$, $\sigma = 10.37$) were recruited. The mean of the approval rate for Swedish participants was 99.39 with a minimum of 94 and a mean of 80.19 studies performed successfully.

Materials

To determine the values associated with the variables of pleasure and arousal
in a target language, we designed an online questionnaire composed of ten
prompts, one for each emotion label. Each prompt followed the same struc-
ture: the title of the prompt contained only an emotion label; then, two 7-item
Likert scales were used to assign the levels of pleasure and arousal associated
with the emotion label. The arousal scale comprised values from *Very sleepy*
to *Very aroused* and the pleasure scale comprised values from *Very miserable*
to *Very pleased*. Note that, in both scales, the value 4 corresponded to the
neutral value. In addition, an attention check question was shuffled along with
the ten prompts of the experiment. In this attention check question, partic-
ipants were asked to assign to the label "Warm" a value between *very cold*
and *very warm* using a 7-item Likert scale as in the rest of the questionnaire
prompts.

Procedure

As in the first experiment, participants had to enter their Prolific ID to regis-
ter for the questionnaire. At the beginning of the experiment, an explanation
of the objective of the experiment together with a brief introduction to the
dimensions of pleasure and arousal was given to the participants. In addition,
instructions describing the task to be performed were also provided. Partic-
ipants were then asked to complete the ten questions of the experiment and
the attention check question with no time limit.

3.5 Results

First experiment

To measure the suitability of a word w to represent the translation of an
affective category c into a target language, we use the *degree of acceptance*.
The *degree of acceptance* ($accept_c(w)$) is defined as the difference between
the number of times the word w has been selected as representative for the
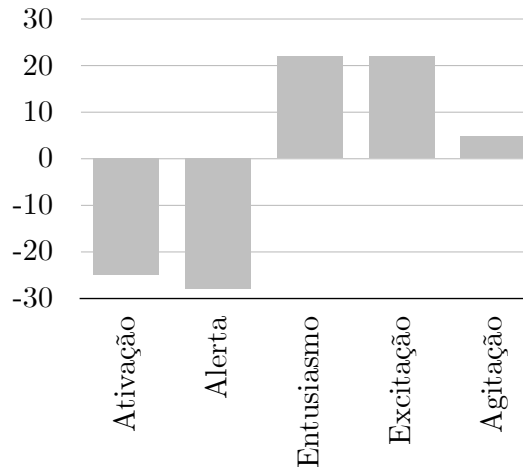
(a) Arousal
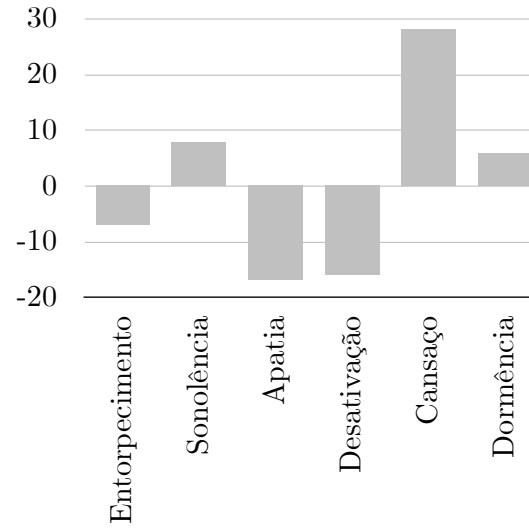
(b) Sleepiness

(c) Pleasure
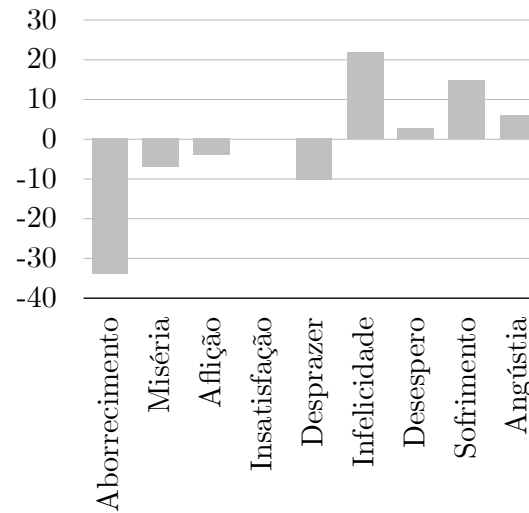
(d) Misery

Figure 3.2: Degree of acceptance of the candidate words in Portuguese for each affective category.

affective category c and the number of times the word has been selected as
non-representative of that affective category.

Portuguese experiment

Figure 3.2 shows the *degree of acceptance* of the candidate words for the four
affective categories. For example, in the case of the affective category *misery*,
the degree of acceptance for "Infelicidade" was $accept_{misery}(Infelicidade) =$
22 indicating that the participants considered this word as very representative
for that affective category. In contrast, the degree of acceptance for "Aborrec-
imento" was $accept_{misery}(Aborrecimento) = -34$ indicating that the majority
of the participants rejected this word to represent the affective category of the
group. Based on the participants' choices in each affective category (see Fig-
ure 3.2), we have obtained the following words with a positive *degree of accep-
tance*: for *pleasure* participants chose the words "Felicidade", "Alegria", and
"Satisfação"; for *misery* the words "Infelicidade", "Sofrimento", "Angústia",
and "Desespero" were chosen; for arousal the selected words were "Entusi-
asmo", "Excitação", and "Agitação"; and for *sleepiness* the words chosen
were "Cansaço", "Sonolência", and "Dormência".

In the case of the label *pleasure* (Figure 3.2c), the words "Júbilo", "Con-
tentamento", and "Prazer" had a negative degree of acceptance. For the
label *misery* (Figure 3.2d), "insatisfação" has a neutral degree of acceptance
and "Aflição", "Miséria", "Aborrecimento", and "Desprazer" has a negative
degree of acceptance. The label *arousal* (Figure 3.2a) has two words with a
degree of acceptance "Ativação" and "Alerta". Finally, the label *sleepiness*
(Figure 3.2b) has three words with negative degree of acceptance "Apatia",
"Desativação", and "Entorpecimento".

The labels generally used for the affective category of pleasure in previous
Portuguese studies [48, 143] are "Prazer" for *pleasure* and "Desprazer" for
misery (*displeasure*), which was an expected result of the first experiment in
Portuguese, which could capture a general meaning of *pleasure* and *misery*.
However, these words were not chosen by the participants. Intuitively, this
result is due to the less intensive meaning of the word "Prazer" in comparison
to "Felicidade" and "Alegria" (chosen by the participants), and the less in-

tensive meaning of the word "Desprazer" in comparison to "Infelicidade" and "Sofrimento" (chosen by the participants). From the experiment results, and by being strict to our unbiased approach, it would not be numerically possible to include the word "Prazer" in Figure 3.2c and the word "Desprazer" in Figure 3.2d because these two words have a negative degree of acceptance by the participants.

Other labels chose by the participants, such as "Sonolência", "Cansaço", "Entusiasmo", and "Contente", are shared with the Portuguese labels of [143], which can suggest their suitability. However, the labels "Dormência", "Agitação", "Desespero" where not included in that study. Intuitively, this result is due to the more intensive or clear meanings of these labels than some of their synonyms, e.g., "Preguiçoso", "Tenso", "Chateado", respectively, included in the study by [143].

In order to perform the second experiment, we decided to select for the affective categories *arousal*, *sleepiness*, and *pleasure*, those words with a positive degree of acceptance, obtaining three words for each affective category. However, by applying the same criteria for *misery*, four possible words were obtained. To avoid bias and misunderstandings, we decided to maintain a homogeneous number of words for each affective category and, therefore, for the affective category *misery* we selected the three words with the highest level of acceptance.

Swedish experiment

The results obtained for the first experiment conducted with Swedish participants are shown in Figure 3.3. Analyzing the *degree of acceptance* of these results, we can appreciate that there is a larger set of words with positive *degree of acceptance* than in the case of the experiment conducted with Portuguese participants. However, it is important that none of the sets have terms that may be contradictory or confusing. If this happens we have two options: we can select three words instead of two (to increase the context) or we can select only one word (to reduce confusion). We ended up using two words for each end-point, which we believe resulted in suitable sets of labels (with sufficient contextual space and minimal confusion) for the second exper-
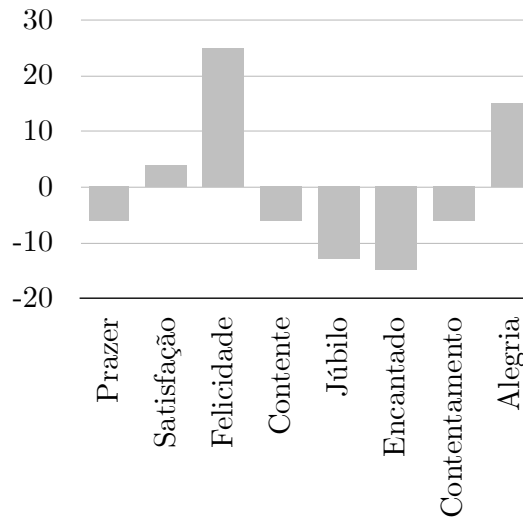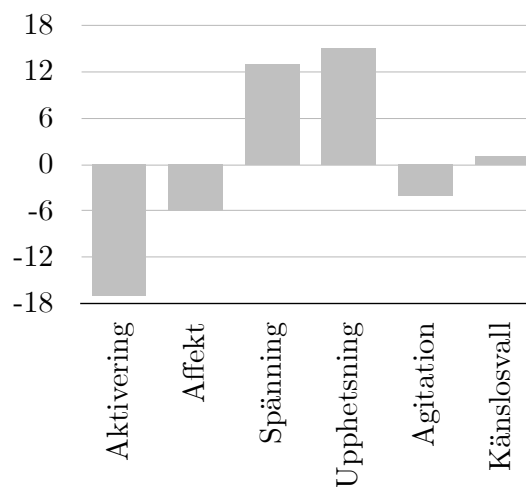
(a) Arousal
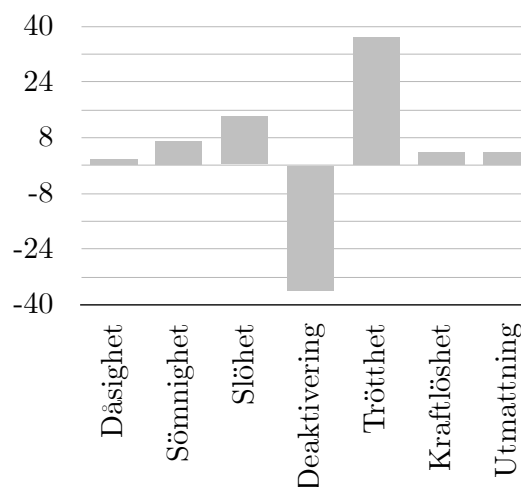
(b) Sleepiness
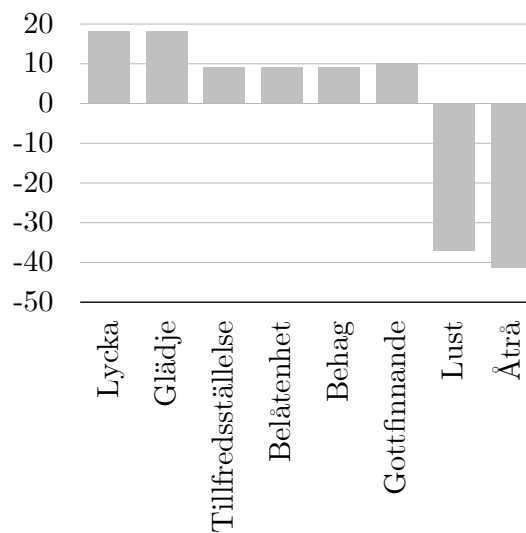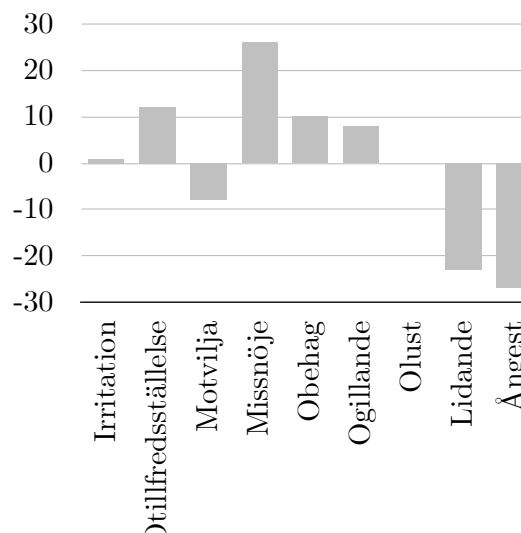
(c) Pleasure

(d) Misery

Figure 3.3: Degree of acceptance of the candidate words in Swedish for each
affective category.

iment. Therefore, based on the participants' choices, this resulted in "Lycka" and "Glädje" for *pleasure*, "Otillfredsställelse" and "Missnöje" for *misery*, "Spänning" and "Upphetsning" for *arousal*, and "Slöhet" and "Trötthet" for *sleepiness*.

Commonly used labels for hedonic tone (*pleasure/misery*) in prior Swedish studies and literature [211, 120] are "Behag" for *pleasure* and "Obehag" for *misery* (*displeasure*). This was an expected result of the first Swedish experiment, which arguably would capture a general meaning of *pleasure* and *misery*. However, these words were not chosen by the participants. Intuitively, this result is due to the *less* intensive meaning of the word "Behag" in comparison to "Lycka" and "Glädje" (chosen by the participants), and the *less* intensive meaning of the word "Obehag" in comparison to "Otill-fredsställelse" and "Missnöje" (chosen by the participants). From the results of the experiment, and by being strict to our unbiased approach, it would not be numerically possible to include the word "Behag" in Figure 3.3c and the word "Obehag" in Figure 3.3d without also including words in other sets that could increase confusion. This result reinforced our decision using two words for each end-point.

Other labels chosen by the participants, such as "Lycka", "Glädje", "Slöhet", and "Trötthet", are common with the Swedish labels of Knez and Hygge [127], which can suggest their suitability. However, the labels "Missnöje", "Spänning", "Upphetsning" where not included in the study by Knez and Hygge. Intuitively, this result is due to the *more* intensive or clear meanings of these labels than some of their synonyms, e.g., "Besvärad", "Upprymd" and "Begeistrad", respectively, included in the study by Knez and Hygge. More intensive and clear labels for end-points have been observed to have clearer conceptions [15], which can suggest their suitability to be used as pleasure and arousal end-points.

Second experiment

To analyse the results of this second experiment, we will refer only to the dimensions of *Pleasure* and *Arousal*, considering that *Misery* is negative *Pleasure* and *Sleepiness* is negative *Arousal*.

Table 3.1: Circular mean and circular standard deviation in degress for each emotion in Portuguese, Swedish, and Spanish languages.

Emotions in English	Emotions in Portuguese			Emotions in Swedish			Emotions in Spanish		
Label/word	Label/word	Mean	Sd	Label/word	Mean	Sd	Label/word	Mean	Sd
Fear	Medo	160.12	39.25	Rädsla	136.93	21.62	Miedo	125.51	15.61
Surprise	Surpresa	62.79	21.02	Överraskad	75.06	18.86	Sorpresa	71.63	26.38
Disgust	Aversão	195.37	41.88	Avsky	167.88	34.00	Asco	182.58	43.65
Anger	Raiva	151.60	37.72	Ilska	140.95	8.20	Enfado	138.55	16.90
Boredom	Aborrecimento	245.96	21.11	Uttråkad	230.33	9.46	Aburrimiento	245.34	21.41
Sleepiness	Sonolência	259.19	18.04	Trött	254.39	16.79	Somnolencia	263.59	15.56
Sadness	Tristeza	207.78	15.51	Ledsen	205.69	20.31	Tristeza	196.02	22.48
Excitement	Excitação	53.18	8.64	Spänning	76.78	20.96	Emoción	39.97	10.32
Calm	Calmo	312.68	46.76	Lugn	325.44	37.77	Calma	318.12	35.89
Happiness	Felicidade	40.22	7.66	Lycka	32.66	15.64	Felicidad	25.09	19.02

In a previous analysis of the results, those responses in which the participants selected both neutral values (i.e., the 4th item in both Likert scales) in the pleasure and arousal dimensions were removed. This decision was made according to previous studies in which it was evidenced that, at values close to zero in both dimensions, the intensity of emotions is so low that it can be considered that there is no emotion [183]. Then, we scaled the obtained responses so that the 1st item of the Likert scale corresponds to the value -1 and the 7th item of the Likert scale to the value 1 of the coordinate axis. Note that the 4th item in the Likert scale corresponds to the value 0 (i.e. the origin of coordinates of the polar representation). Once scaled, we calculated the angle of each response from the pleasure and arousal values. This angle represents an approximation to the conceptualization that each participant associates with an emotion in the pleasure and arousal space. Finally, we estimated the dispersion of the samples for each emotion in each of the two languages and we removed the possible outliers.

Table 3.1 summarizes the results obtained in the second experiment in the Portuguese and Swedish languages, as well as the results of the previous experiment conducted with Spanish-speaking European participants presented in [220]. This table presents the mean angle for each emotion together with its circular standard deviation obtained using the Formula 3.1. To facilitate the visualization and the analysis of the distribution of the ten emotions in the circular space composed of the pleasure and arousal dimensions, we proposed

the use of the circular boxplots [36] that can be seen in Figures 3.4 and 3.5 for Portuguese and Swedish respectively.

Portuguese experiment

The values of pleasure and arousal associated with the 10 emotions in Portuguese resulting from the second experiment can be seen in Table 3.1.

In this experiment, some emotions had a much larger standard deviation than others. It can be a result of either unclear labelling or that the participants had a clearer conception of some emotions, making it easier to define them in terms of pleasure and arousal. For instance, "Excitement" and "Happiness" had more clear responses from the participants, while emotions such as "Calm" and "Fear" seemed to be more challenging to define. This goes in line with previous research, which has observed that stronger/intense emotions, such as "Happiness", have more precise conceptions (emotional differentiation). Similar observations could be made when comparing languages. For some emotions, the standard deviation had apparent differences between the Portuguese test and the Swedish test. The Portuguese test had much clearer "Excitement", "Happiness", and "Sleepiness" definitions. Intuitively, it can be that these emotion labels have more precise definitions in Portuguese. Comparing the Swedish, Spanish, and Portuguese emotion measures, we can observe similarities on "Excitement", "Boredom", and "Sleepiness" on Portuguese and Spanish measures (Table 3.4).

Based on Table 3.1 and Figures 3.4 and 3.7, we can observe: i) positive arousal and positive pleasure are related to the emotions "Happiness", "Excitement", and "Surprise"; ii) positive arousal and negative pleasure are related to "Anger" and "Fear"; iii) negative arousal and negative pleasure are related to "Disgust", "Boredom", "Sleepiness"; and iv) negative arousal and positive pleasure are related to "Calm".

Furthermore, we can verify that they keep their values in the quadrants, considering the median and the standard deviations "Happiness" and "Excitement" (positive arousal and pleasure). For the other emotions, and based on Figure 3.4, 75% of the results belong to the quadrant of the median of each emotion, but 25% of the results belong to different quadrants.

(a) Fear

(b) Anger

(c) Disgust

(d) Sadness

(e) Boredom

(f) Sleepiness

(g) Calm

(h) Happiness

(i) Excitement

(j) Surprise
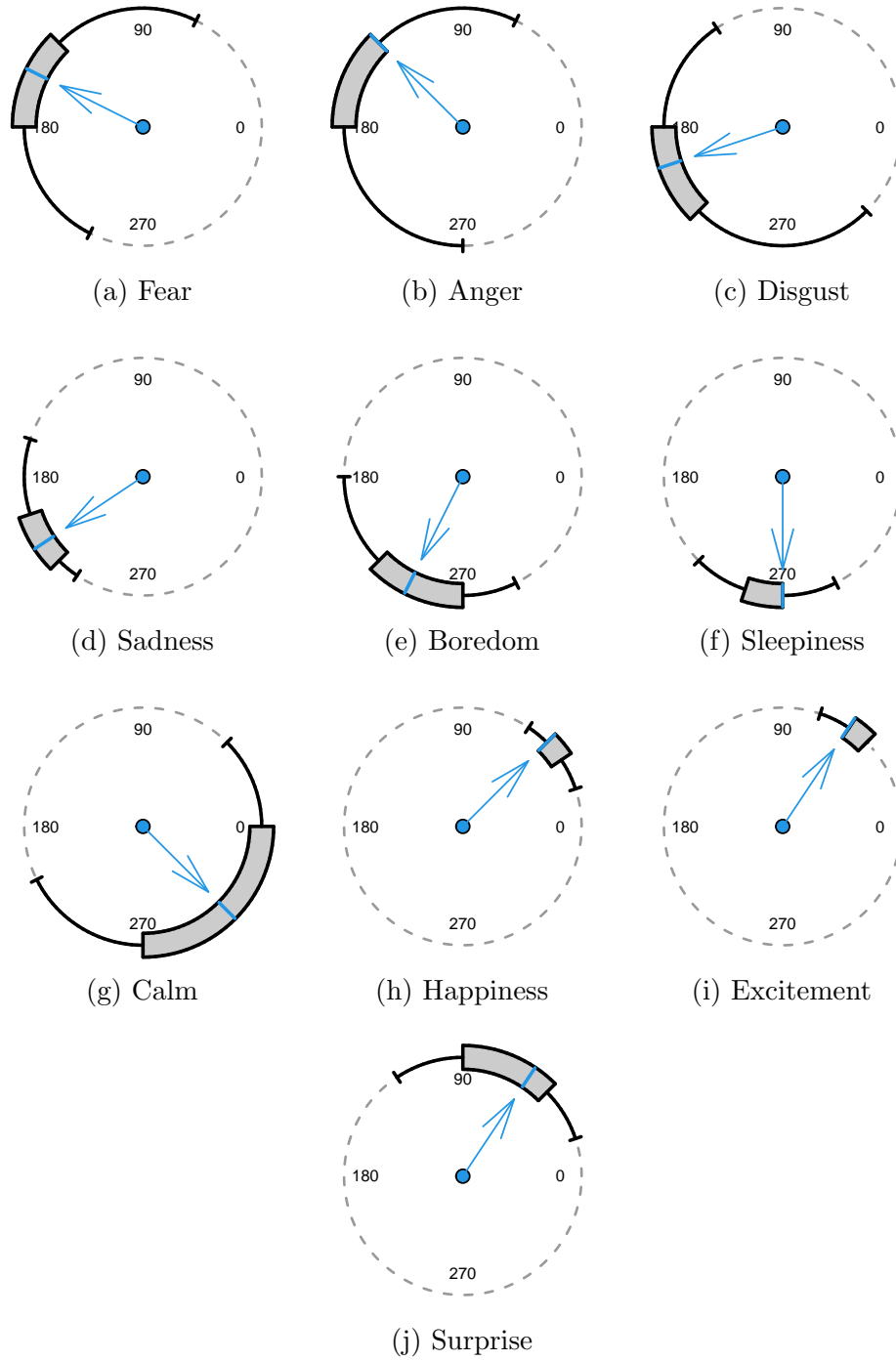
Figure 3.4: Portugal experiment results.

According [222], the proximity or distance between the emotions represented circumference assume the similarity or difference between emotions. The model defines that emotions are less positively correlated when they are approximately 90 degrees apart. At 90 degrees apart, two affective states should be very little or not correlated. In turn, at 180 degrees of distance, affective states must be negatively correlated. Thus, based on Figure 3.7a), we can verify that "Sleepiness"/"Boredom" is negatively correlated with "Excitement"/"Happiness".

Swedish experiment

Similar to the Portuguese experiment, in the second Swedish experiment, some emotions had a much larger standard deviation than others. Once again, this can be a result of either unclear labelling or that the participants had a more clear conception about some emotions, making it easier to define them in terms of pleasure and arousal. For instance, "Anger" and "Boredom" had more clear responses from the participants, while emotions such as "disgust" and "Calm" seemed to be more difficult to define. This also goes in line with previous research, which has observed that stronger/intense emotions, such as "Anger", have clearer conceptions (emotional differentiation) [15].

Similar observations could be made when comparing languages. For some emotions, the standard deviation had clear differences between the Swedish test and the Portuguese test. A reason for this can be cultural as well as personal. For instance, "Excitement" (translated to "Spänning" in Swedish) that in English commonly is perceived as a positive emotion, can in Swedish (depending on the context) be perceived as either negative (e.g., too exciting or a scary level of thrill) or positive (e.g., having fun or expecting/waiting for something good). This is observed in the Swedish test result, where "Excitement" spans from pleasure to displeasure, although pleasure is dominant (see Figure 3.6). The Portuguese test, however, had much clearer "Excitement" definitions, being strictly pleasurable (see Figure 3.6). Other emotions, such as "Boredom" and "Anger" had much smaller standard deviations in the Swedish test compared to the Portuguese test (see Figure 3.4 and 3.5). Intuitively, it can be that these emotion labels have more precise definitions in

(a) Fear

(b) Anger

(c) Disgust

(d) Sadness

(e) Boredom

(f) Sleepiness

(g) Calm

(h) Happiness

(i) Excitement

(j) Surprise
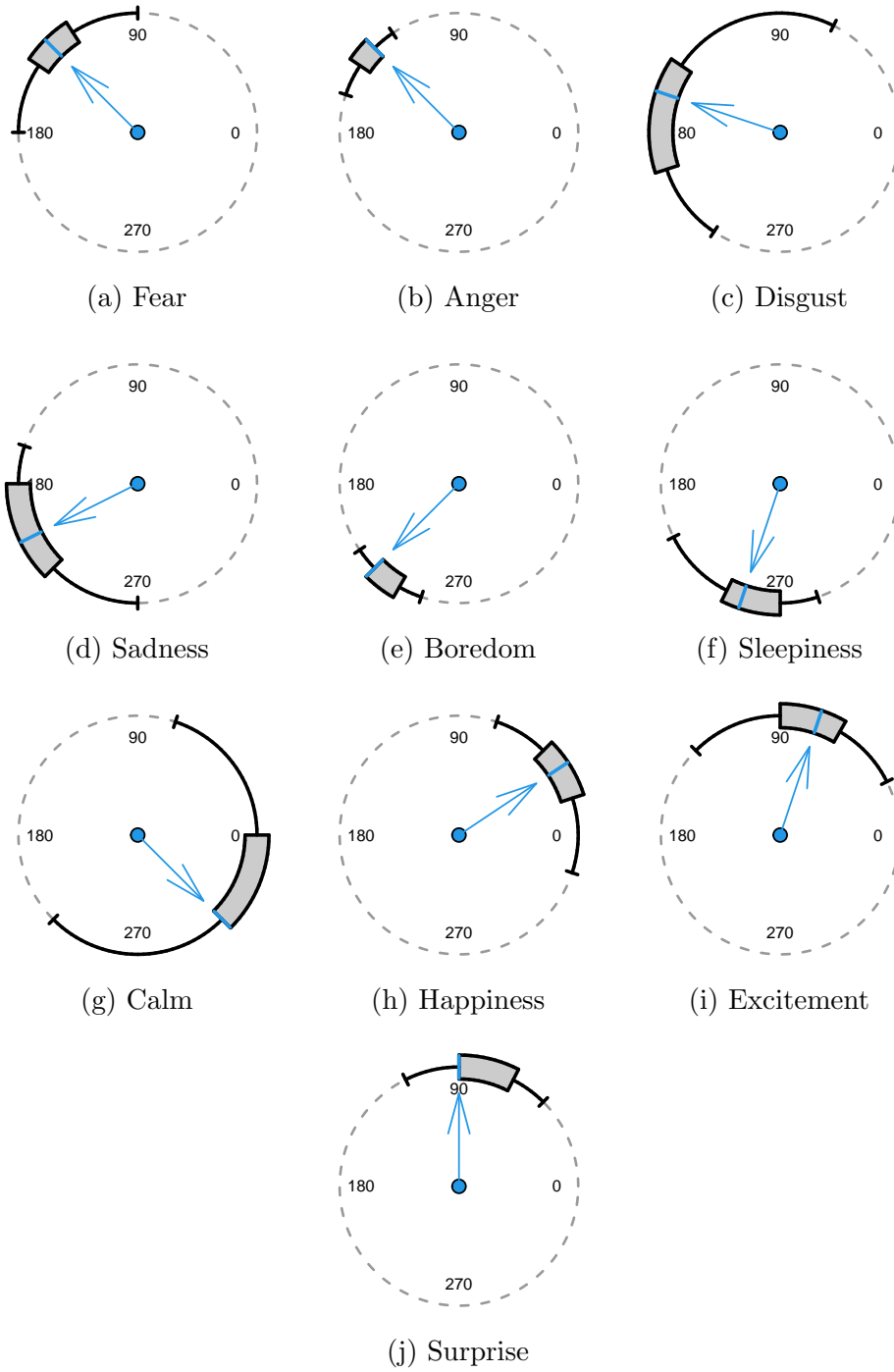
Figure 3.5: Sweden experiment results.

Swedish.

3.6 Discussion

Comparison between languages

Analyzing the results of the first experiment shown in Figure 3.2, the Portuguese label "Prazer" (pleasure) has a negative impact, which may be explained based on the cultural and religious aspect, because on Portuguese meaning, "Prazer" (pleasure) is associated to sexual pleasure instead of happiness and have a negative opinion on society. In the same way, the label "Desprazer" (displeasure) also have a negative impact on the label choice, which sustain the previous explanation.

From the second experiment, comparing the Portuguese and Swedish emotion measures (see Figure 3.4 and 3.5), we can observe a small decrease in arousal and pleasure for the Portuguese on all "negative" emotions (i.e., "Fear", "Anger", "Disgust", "Sadness", "Boredom", and "Sleepiness"), approximately increasing the angle representation of each Portuguese emotion with 5-10 degrees. While the same could not be observed for more "positive" emotions (i.e., "Excitement", "Happiness", and "Surprise"), an interesting point is that the difference is similar/consistent for all emotions in set "negative" emotions. Furthermore, in Figure 3.7, we can observe that "Fear" and "Anger" have less intense arousal for Portuguese results on the second experiment than Spanish and Swedish results. Also, for "Happiness", the Portuguese result is more intense in terms of pleasure and arousal. Finally, the "Calm" emotion has less intense pleasure than Spanish and Swedish, but the arousal intensity is similar to the Swedish results.

From these observations we may conclude that: i) there is a cultural difference in the subjective experience of these emotions, and ii) there seems to be a cultural modifier that is consistent across emotions. We may thus make a third conclusion: iii) we are measuring similar emotions in the Swedish experiment and the Portuguese experiment, if normalized with the observed cultural modifier. This is interesting from a modelling perspective. If able

(a) Portuguese
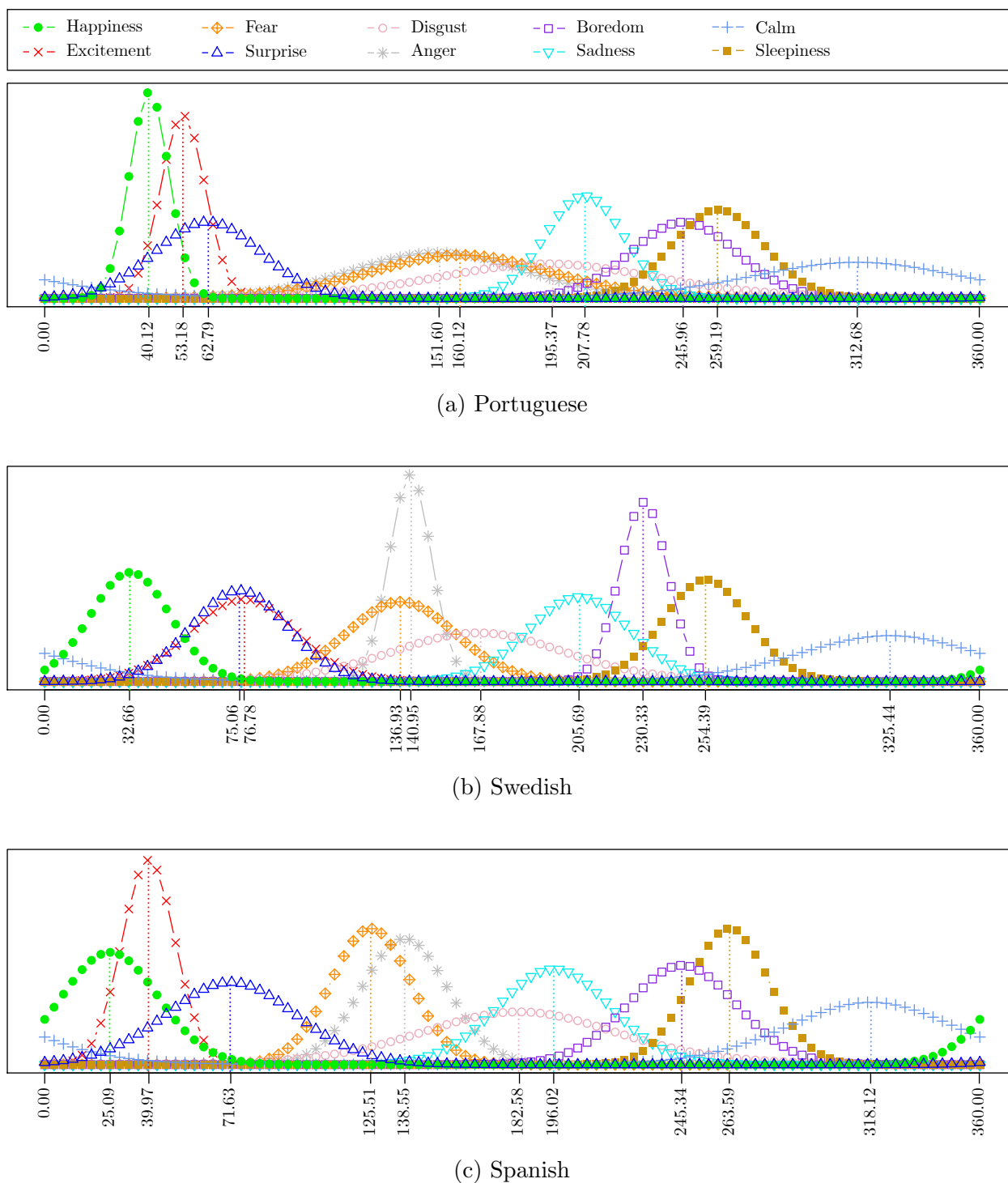
(b) Swedish

(c) Spanish

Figure 3.6: Mixture of Von Mises curves of the ten emotions in Portuguese, Swedish, and Spanish. The axis represents the angles α_e in degrees.
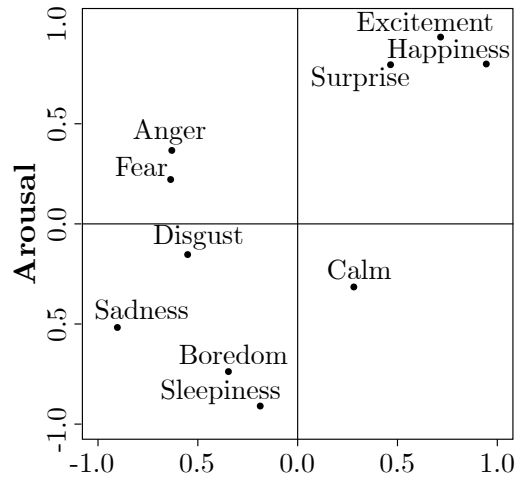
(a) Portuguese
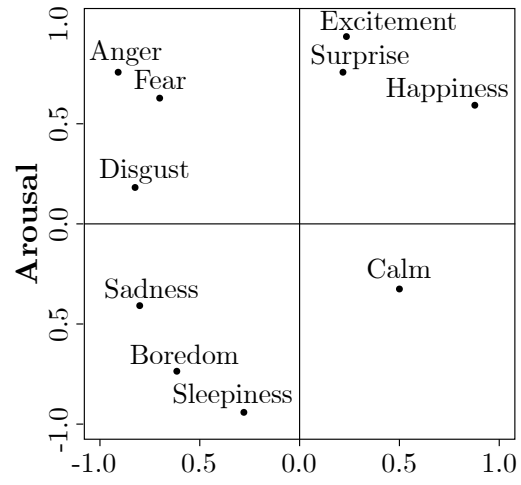
(b) Swedish
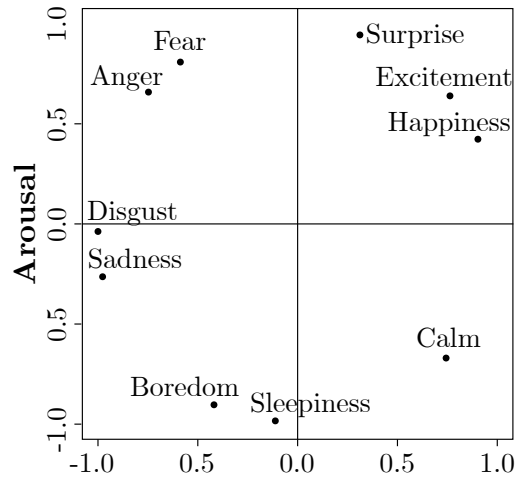
(c) Spanish

Figure 3.7: Results obtained.

Table 3.2: Results of the chi-square test for the comparison of the distribution of emotions across languages. Note that * indicates p-value < 0.05.

Emotion	Portuguese - Swedish		Portuguese - Spanish		Swedish - Spanish	
	χ^2 value (df)	p-value	χ^2 value (df)	p-value	χ^2 value (df)	p-value
Fear	45.899 (15)	0.000*	75.95 (15)	0.000*	23.951 (8)	0.002*
Surprise	21.884 (10)	0.015*	30.313 (12)	0.002*	23.119 (11)	0.017*
Disgust	53.519 (18)	0.000*	19.224 (18)	0.378	56.231 (18)	0.000*
Anger	54.674 (14)	0.000*	25.506 (14)	0.029*	31.708 (7)	0.000*
Boredom	56.461 (8)	0.000*	3.314 (8)	0.913	58.382 (9)	0.000*
Sleepiness	8.8291 (7)	0.265	12.584 (7)	0.082	21.064 (8)	0.006*
Sadness	18.458 (9)	0.030*	22.028 (11)	0.024*	24.191 (13)	0.029*
Excitement	83.071 (9)	0.000*	61.470 (7)	0.000*	121.82 (12)	0.000*
Calm	26.338 (15)	0.034*	16.788 (12)	0.157	12.567 (14)	0.560
Happiness	22.512 (8)	0.004*	43.523 (7)	0.000*	20.367 (9)	0.015*

to measure and identify cultural modifiers for the experience of emotions (independent of particular emotions), these can work as heuristics in shared representations of emotions across cultures.

We performed a comparative analysis of the Portuguese, Swedish, and Spanish samples to identify the existence of possible discrepancies between languages when assessing the levels of pleasure and arousal for the ten emotions. Since the nature of the experiment implies that we have a finite number of possible combinations of responses and, consequently, we have a finite number of possible angles. Therefore, we decided to apply a Chi-square test to test whether there were significant differences between language and the angles associated with emotions. The starting hypotheses for this analysis are: the null hypothesis h_0 "the angle associated with an emotion is independent of culture and language"; and the alternative hypothesis h_1 "the angle associated with an emotion and culture and language have some degree of dependence".

Table 3.2 shows the results of this analysis. When testing whether there is a dependence between Portuguese and Swedish we can appreciate that, except for the case of "Sleepiness", we can reject the null hypothesis h_0 with a confidence level of 95%, i. e., there is a language dependence in the selection

of emotion angles. In other words, for nine of the ten emotions, the differences between the emotion angles are statistically significant between languages. The same applies when comparing Swedish with Spanish. As we can see, there is statistically significant evidence that the selection of the emotion angle depends on language in nine of the ten emotions (all except "Calm"). However, when comparing Portuguese with Spanish, we can observe that these differences are reduced. In this case, we find four emotions ("Disgust", "Boredom", "Sleepiness", and "Calm") for which we do not have enough evidence to reject the null hypothesis h_0 with a confidence level of 95%. These results are interesting, since at the linguistic and cultural level, there are greater similarities between Portuguese and Spanish than between Portuguese and Swedish or Spanish and Swedish. Considering the fact that Portuguese and Spanish share a greater lexical similarity since both are included in the *Ibero-Romance* language family group [168], these results are in line with the results obtained in previous work with the study developed in [107] (described in Section 3.2) in which similarities between language families were observed.

A cross-cultural model of affect adapted to Portuguese and Swedish

As discussed in Section 3.3, although a dimensional model provides a general representation of affective states, it is necessary to adapt the representation of affective labels to the language in which the model is going to be used. For this purpose, from the results obtained in the experiments with Portuguese and Swedish participants, we estimated the probability regions for the ten emotions in the pleasure and arousal space using the Von Mises distribution described in Formula 3.3. Figure 3.6 shows the estimated Von Mises distributions for the ten emotions in Portuguese, Swedish, and Spanish. From these distributions, we have generated an emotional representation model for the Portuguese and Swedish languages following the model presented in Section 3.3 for the Spanish language. Figures 3.8 and 3.9 show a graphical representation of the probability regions for the Portuguese and Swedish affect representation models using one standard deviation. As in Figure 3.1, the outer rings of Figures 3.8 and 3.9 represent the regions assigned to each emo-
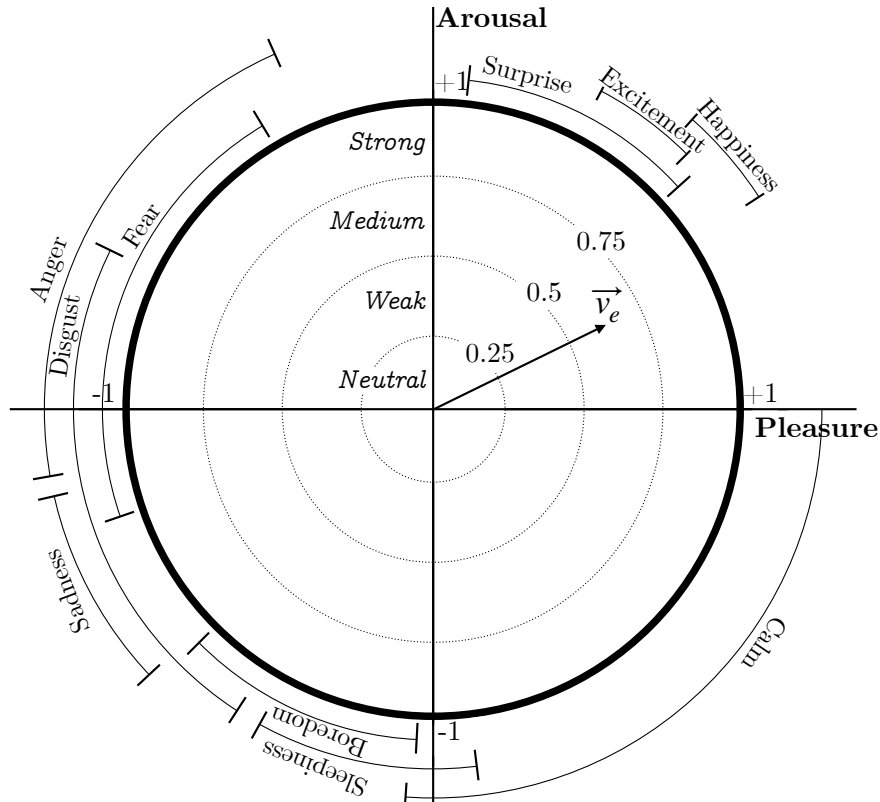
Figure 3.8: Distribution of probability functions in the pleasure-arousal space
using one standard deviation for Portuguese.

Table 3.3: Example of the results of the *Emotion Representation* process for
the emotion "Happiness" with a *Medium* intensity.

	α	$|v|$	Pleasure	Arousal
Portuguese	40.22°	0.625	0.477	0.404
Swedish	32.66°	0.625	0.526	0.337
Spanish	25.09°	0.625	0.566	0.265

tion in each language, while the inner rings represent the different levels of
intensity.

Once defined the representation models of affect for Portuguese and Swedish,
to allow the representation, interpretation, and expression of affective states
using a polar representation, we designed two processes (described in detail in
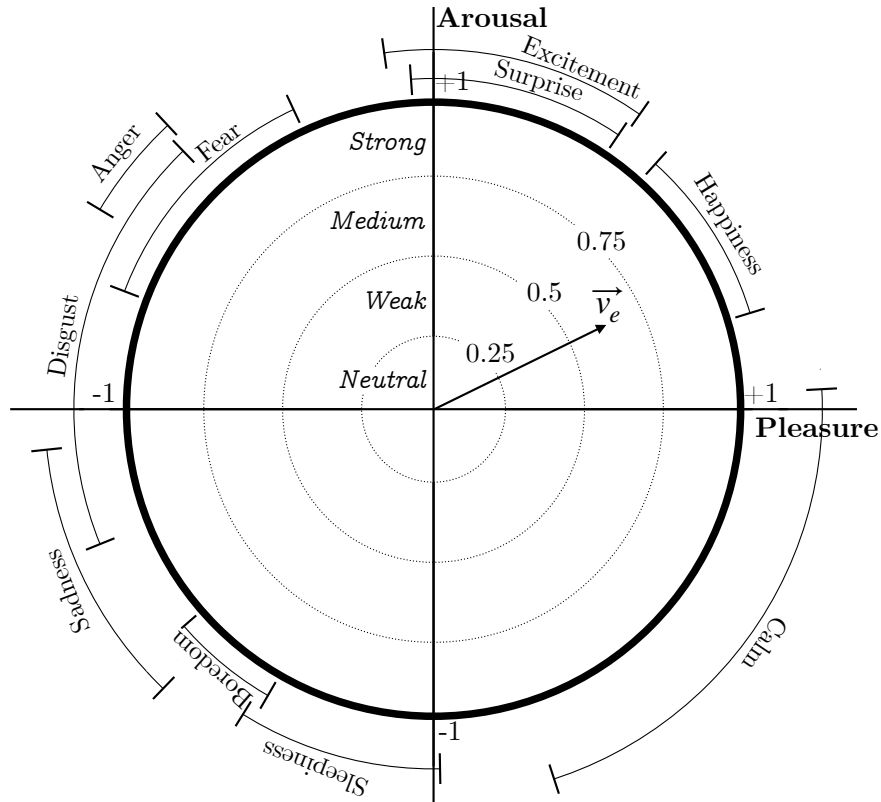
Figure 3.9: Distribution of probability functions in the pleasure-arousal space using one standard deviation for Swedish.

[220]): the *Emotion Representation* process and the *Emotion Expression* process. The *Emotion Representation* process is a defuzzification process that, based on the mean values obtained in the experiments for each emotion, allows the agent to represent in the pleasure-arousal space the emotions recognized or internally elicited as part of an affective process (e.g. appraisal process). For example, when an affective agent needs to represent the emotion "Happiness" with a *Medium* intensity, the *Emotion Representation* process uses the mean angle α of the emotion "Happiness" adapted to the cultural environment to estimate the angle of the vector α. To estimate the modulus of the vector $|v|$ we use the midpoint of each intensity region. In the example, the *Medium* intensity corresponds to 0.625 (the midpoint between 0.5 and 0.75) (see Figure 3.8). Table 3.3 shows the values of pleasure and arousal when applying the *Emotion Representation* process to emotion "Happiness" with

Table 3.4: $\mu_{A^i}(u)$ values for the example ($u = 0.625$).

	Neutral	Weak	Medium	Strong
$\mu_{A^i}(0.625)$	0	0	1	0

Table 3.5: Example of $\mu_{B^j}(\alpha)$ values for $\alpha = 245°$. For the Portuguese and Swedish labels see Table 3.1.

	Fear	Surprise	Disgust	Anger	Boredom	Sleepiness	Sadness	Excitement	Calm	Happiness
Portuguese	0.075	≈ 0	0.374	0.037	0.998	0.723	0.055	≈ 0	0.253	≈ 0
Swedish	≈ 0	≈ 0	0.059	≈ 0	0.311	0.844	0.147	≈ 0	0.079	≈ 0
Spanish	≈ 0	≈ 0	0.271	≈ 0	0.999	0.479	0.087	≈ 0	0.102	≈ 0

a *Medium* intensity in Portuguese, Swedish, and Spanish. The formulas to obtain these values are described in detail in [220].

The *Emotion Expression* process is a fuzzification process that allows an affective agent to express an affective state using the same label/word in the same language as its human interlocutor. We provide two fuzzy functions, $\mu_{A^i}(u)$ and $\mu_{B^j}(\alpha)$ (described in detail in [220]), to estimate the most appropriate terms to express the intensity and type of emotion, respectively, according to the cultural environment. To illustrate how these fuzzy functions work, let us consider the following example in which an affective agent wants to express its affective state represented by the vector $\vec{v}_e$ (see Figure 3.8), formed by the components *Pleasure* $= -0.264$ and *Arousal* $= -0.566$. The angle of the vector is $245°$ and the modulus of the vector is 0.625. According to the modulus of the vector, the formula for the intensity of the *Emotion Expression* process produces the results shown in Table 3.4. Therefore, the intensity level of the emotion is *Medium*. Table 3.5 shows the estimated probability of the possible emotion labels for the Portuguese, Swedish, and Spanish languages applying the formula for the type of the *Emotion Expression* process to a direction of the vector of $245°$. According to this results, we can conclude that the most probable emotion label to be used by the agent for the Portuguese and Spanish languages is "Boredom" (i.e., "Aborrecimento" for Portuguese and "Aburrimiento" for Spanish) with a probability of 0.998 and 0.999 respectively, while for Swedish it is "Sleepiness" (i.e., "Trött") with a probability of 0.844.

Table 3.6: Results of the chi-square test for the study of the influence of gender and age on the assignment of pleasure and arousal values for the ten emotions in the Portuguese and Swedish samples. Note that * indicates p-value < 0.05.

| | Portuguese | | | | Swedish | | | |
| | Gender | | Age | | Gender | | Age | |
Emotion	χ^2 value (df)	p-value	χ^2 value (df)	p-value	χ^2 value (df)	p-value	χ^2 value (df)	p-value
Fear	11.084 (14)	0.679	12.471 (14)	0.568	6.789 (8)	0.559	5.618 (8)	0.056
Surprise	9.055 (8)	0.337	18.130 (8)	0.020*	5.618 (6)	0.467	16.206 (6)	0.766
Disgust	18.528 (14)	0.183	12.134 (14)	0.595	16.206 (13)	0.238	3.520 (13)	0.382
Anger	15.896 (14)	0.319	18.255 (14)	0.195	3.520 (4)	0.474	1.346 (4)	0.133
Boredom	8.443 (7)	0.295	4.610 (7)	0.707	1.346 (4)	0.853	10.313 (4)	0.130
Sleepiness	9.380 (6)	0.153	9.619 (6)	0.141	10.313 (6)	0.112	10.059 (6)	0.092
Sadness	6.808 (6)	0.338	9.303 (6)	0.157	10.059 (8)	0.260	11.437 (8)	0.043*
Excitement	1.656 (3)	0.646	5.300 (3)	0.151	11.437 (9)	0.246	10.237 (9)	0.559
Calm	3.954 (9)	0.914	15.151 (9)	0.086	10.237 (14)	0.744	6.179 (14)	0.152
Happiness	2.893 (4)	0.575	3.502 (4)	0.477	6.179 (7)	0.518	0.518 (7)	0.054

Gender and age comparison

As in the case of language, we performed a comparative analysis of the samples to identify the existence of possible discrepancies between gender and age when assessing the levels of pleasure and arousal for the ten emotions using Chi-square test. Applying a criterion of homogeneity for sample size, we divided the age into two groups according to the median, which was 29 for the Swedish participants and 32 for the Portuguese participants.

Table 3.6 shows the results of the comparative analysis of the samples for gender and age using the Chi-square test. The starting hypotheses for this analysis are: the null hypothesis h_0 "the angle associated with an emotion is independent of gender/age"; and the alternative hypothesis h_1 "the angle associated with an emotion and the gender/age have some degree of dependence". Regarding gender, we can see that in the experiment conducted with both Portuguese and Swedish participants, the *p-value* of the results indicate that there is not enough evidence to reject the null hypothesis h_0 at a confidence level of 95%. Therefore, there is not enough evidence to determine that there is a dependence between gender and the angle that participants associated with emotions.

Similarly, attending to the general results for the age comparison, we can

observe that in both languages there is not enough evidence to indicate that there is a dependence between age and the angle associated with emotions. However, we find two emotions, surprise in the case of Portuguese and sadness in the case of Swedish, where the *p-value* is less than 0.05. That is, with a confidence level of 95%, the null hypothesis h_0 is rejected. In other words, there seems to be a dependence between the selected angle and age for these two emotions. This result may be due to a bias in age that occurs when dividing the samples into two homogeneous groups comprising different age ranges.

Importance of multidimensional models

We can clearly observe, by the results of this study, that emotions, defined by the dimensions of pleasure and arousal, do not mean the same thing everywhere. To design universal affective agents, we need multidimensional models that capture these cultural differences. People from different parts of the world can physically and psychologically feel emotions in a similar way, but perceive, interpret, or convey the emotions differently. Such differences in perception of emotions can contribute to what attitude people have toward certain emotions and emotional expressions. Affective agents must be able to represent, detect, and express emotions in a way that follows the individual's conceptions. Natural language processing (NLP) has mainly been based on supervised learning, trained on large sets of conversational data, where deep learning methods have seen great potential. In the context of emotion detection NLP is usually based on finding specific predefined keywords as "Surprised", "Sad", and "Angry" [84]. The observed cultural differences in the interpretation of emotions can thus be miss represented in these models. The multidimensional model presented in this study can provide affective agents with enhanced capability of emotional differentiation, which is increasingly important in phase with the progressed development of NLP-based conversational interfaces.

3.7 Applications

A cross-cultural affective agent model

Although the representation based on the dimensions of pleasure and arousal can be considered universal in the definition of the agent's internal structure of knowledge representation, for an affective agent to be able to represent, interpret, and express affective states it is necessary to adapt this representation to the cultural environment in which it is going to be used. In addition, when an affective agent evolves in several different cultural environments, it also requires the ability to adapt its affective behavior to these environments. This can be achieved by defining a specific pleasure and arousal space for each cultural environment. In this paper, we present three different models of affect representation adapted to Portuguese, Swedish, and Spanish. All these two-dimensional representations can be easily incorporated into the affective component of a multicultural affective agent allowing it to interpret, represent, and express affective states using the same language as its human interlocutor and adapting to his/her cultural variations. Then, through the *Emotion Representation* process and the *Emotion Expression* process, the agent can represent and express different affective states in different languages. The example presented in Section 3.6 illustrates how an affective agent could use these models to adapt its behavior to the cultural environment of its human interlocutor.

Application in Human-Aware Planning

The multicultural representation model of affect proposed in this paper can be utilized in a human-aware planning scenario [41], where a software agent requires a model of a human user's emotional state for planning its actions in interaction with the user.

Let us consider a software assistant, designed to manage a human user's stress during their interaction by adjusting its behavior or output. The agent utilizes a logic-based knowledge base that captures: i) a model for representing emotions, ii) a human stress-model, in which different sets of emotions
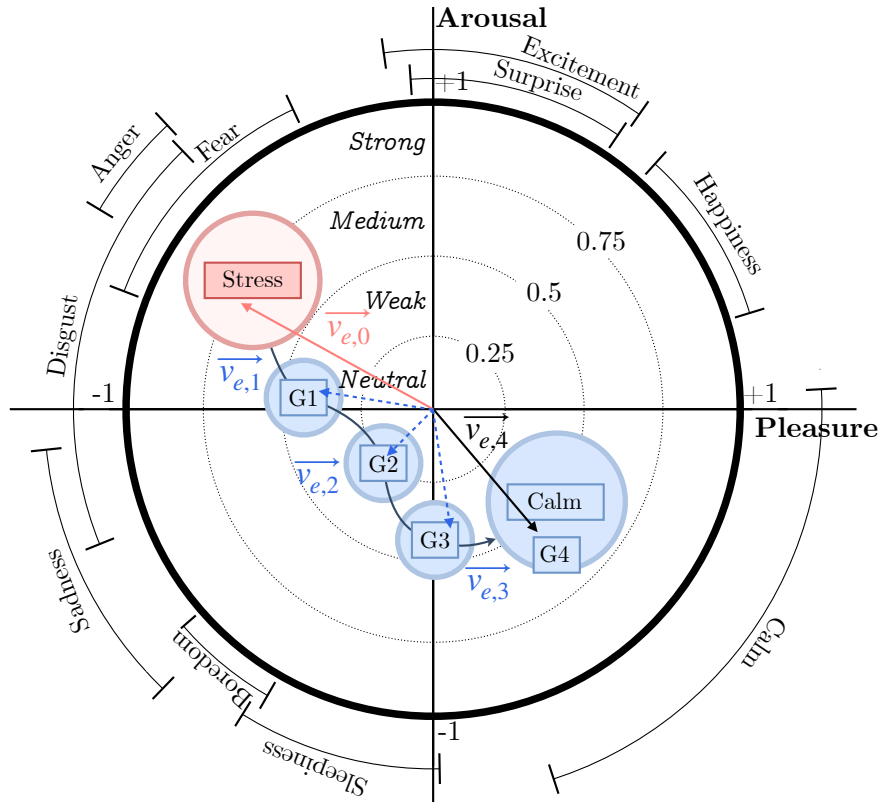
Figure 3.10: A stress relief plan consists of a sequence of subgoals (G1-G3) and a final goal (G4).

are linked to stress-states, and iii) a decision-model, through which the agent generates a sequence of actions aiming to change a set of unwanted emotions to a set of wanted emotions. The agent does this by influencing the user's pleasure and arousal through appropriate interactive actions. In order to identify goals (suitable emotional change) and avoid obstacles (less suitable emotional change), the agent deliberates about alternative plans by considering emotions defined in the pleasure-arousal space (see Figure 3.10).

Stress-model: Previous research has observed that psychological stress-responses often include a complex array of "negative" emotions, such as depression, anxiety, anger, and distress [63]. Following these results, we can define *negative stress* as a set of emotions, which share a high level of arousal and (usually) low pleasure. This would mean that the upper-left part of the pleasure-arousal space captures a set of emotions that represent negative

stress. A desired goal of the agent is to decrease stress. A stress-relief state can (intuitively) be defined by a set of emotions with lower arousal and/or higher pleasure. However, there may be more or less suitable ways for achieving stress relief. Whether to decrease arousal or increase pleasure in a moment depends on the current emotional state of the human, the current context, and a variety of personal and cognitive attributes. Desirable plans for stress relief are generated by the system's decision-model.

Decision-model: A method for defining a decision-model for picking suitable actions to influence emotions is to conduct knowledge elicitation processes [208] with relevant experts, such as psychologists. By consulting experts (as well as relevant psychological theories) we can understand how to appropriately deal with emotional change, and capture this knowledge in explicit decision-rules. These rules helps the agent to anticipate how the mental model of the human will change and pick suitable actions in each state of the interaction. A decision-model can be modelled in the structure of a transition system [83] (consisting of states S and transitions between states), in which each state corresponds to a set of emotions, and transitions correspond to suitable change, triggered by an action or a set of actions, between emotional states. A sequence of actions that trigger transitions between states constitute a plan to reach emotions that represent stress relief (e.g., from a state representing "Stress", a set of actions and transitions lead to a state representing "Calm"). Through the constraints of the transition system, the agent can decide which emotions to aim for next, depending on the current emotional state. The decision-model can, for instance, be implemented in Action Reasoning [83]. In action reasoning, a set of (emotion influencing) actions A are defined. An action $a \in A$ has preconditions, stating from which emotional states an action can be executed, and post-conditions, stating the estimated emotional change (proceeding state) after the action is executed. The solutions of the planning problem (the generated plans) are returned as trajectories of length n, $\langle s_0, A_1, s_1, A_2, \ldots, A_n, s_n \rangle$, which are sequences of sets of actions $\subseteq A$ and states s_n of S satisfying the conditions for suitable emotional change specified by the decision-model.

Interaction: The actions of A are by extension connected to specific interactive actions by the agent, depending on how it is embodied/visual-

ized and on the particular application, e.g., changing its behavior, facial expressions, voice feedback, or by other means influencing the human's beliefs, indirectly changing pleasure and arousal. For instance, interactions could result in the following trajectory, consisting of emotional states and sets of actions: $\langle s_0 : \{Fear, Anger\}$, $A_1 : \{Pause$, Speak slowly,Clarify expectations$\}$, $s_1 : \{Nervousness\}$, $A_2 : \{Smile, Show\ instructions\}$, $s_2 : \{Collected\}$, $A_3 : \{Pause, Smile, Kind\ voice\}$, $s_3 : \{Calm\}\rangle$.

The interaction is dependent on reliable ways to elicit pleasure and arousal levels of the human user. Depending on the application, this can for instance be achieved by using appropriate machine learning models for classifying pleasure and arousal from psychological data, and/or through a user interface, where the human user supplies the system with subjective emotional data in terms of emotion-labels (in the user's own language).

3.8 Conclusions

One of the main challenges in improving human-machine interaction and simulations of human behavior and human social organizations is to establish models that allow agents to represent, express, and interpret affective states. In this sense, dimensional representations of affect seem to be promising since they allow modeling affective states through continuous dimensions. However, it is necessary to consider that, in dimensional approaches, cultural and language variations in the values of the dimensions associated with the different affective states arise. Therefore, it is necessary to establish representation models of affect capable of adapting to these cultural variations. In this paper, we show the results of experiments conducted to adapt a model of affect representation to the Portuguese and Swedish language. For this purpose, we apply a methodology based on two experiments. In the first experiment, we establish the most feasible translations for the affective categories that define the dimensions of pleasure and arousal. The obtained results show the importance of performing this type of experiments when translating complex cognitive constructs such as emotions or affective categories. For example, in the case of Swedish, the most common translation for the pleasure dimen-

sion is "Behag". However, participants preferred to choose the words "Lycka" and "Glädje" while, for misery, participants preferred "Otillfredsställelse" and "Missnöje" over "Obehag" which is the most common translation. The same pattern was observed in Portuguese, where the most common translations in the literature for the pleasure dimension are "Prazer" and "Desprazer". However, participants preferred to choose "Felicidade" and "Alegria" for pleasure and "Infelicidade" and "Sofrimento" for misery.

Based on the results of the first experiment, in the second experiment we explored the regions associated with the different emotions in the pleasure-arousal space. To avoid bias, we decided to contextualize each affective category with the words most rated by participants. The analysis of the results of this second experiment shows the importance of adapting the models of affect representation to the cultures and languages in which they will be used. We have shown the existence of significant differences in the values of pleasure and arousal associated with emotions across languages and cultures. These differences increase when the languages come from different language families, as in the case of Portuguese and Swedish. The results also show that there do not appear to be significant differences in gender or age when associating levels of pleasure and arousal with emotion labels.

The results of the second experiment have been used to design two new computational models of affect representation based on the dimensions of pleasure and arousal adapted to Portuguese and Swedish languages. In both models affective states are represented as vectors in which the direction indicates the type of affective state and the modulus indicates its intensity. To determine the type of affective state we use a representation based on probability regions. Probability regions allow us to estimate, with a certain degree of uncertainty, the probability that a vector belongs to a certain type of affective state.

We propose two applications for these models. On the one hand, we propose a model of affective agent capable of evolving in different cultural environments. The affective agent incorporates multiple models of affect representation adapted to different cultural environments allowing it to adapt its affective behavior according to the language and culture of its human interlocutor. On the other hand, we propose the use of these models in a

human-aware planning scenario. The agent represents the affective state of
the user as a function of his cultural environment. Then, the agent uses
the vector representation to guide its actions in order to influence the user's
affective state and redirect it towards a target affective state.

The multidimensional model presented in this study can provide affec-
tive agents with enhanced capability of emotional differentiation, which is in-
creasingly important in phase with the advances of NLP-based conversational
interfaces. The proposed model can enable affective agents to represent, de-
tect, and express emotions in a way that follows an individual's conceptions,
independent on cultural context.

We are currently working on the development of the application models
proposed in this paper. We are also interested in studying the influence of
other dimensions, such as the dominance dimension, on the model of affect
representation. In particular, we are interested in analyzing whether these
dimensions allow us to improve the identification of affective states such as
"Fear" or "Anger".

A fuzzy appraisal model for affective agents adapted to cultural environments using the Pleasure and Arousal dimensions

Joaquin Taverner, Emilio Vivancos, and Vicente Botti

Abstract

Humans use rather vague and imprecise words to express emotions. Therefore, fuzzy logic allows computational affective models to use emotions in the same way that human beings express them. However, people from different cultures and languages assign different meanings to the same emotion word. Unfortunately, there are still no affective computing models that really take these two factors into consideration. In this paper, we propose a fuzzy model of appraisal for multi-agent systems that is adapted to Spanish-speakers. Our methodology has two steps. First, the agent evaluates an event using a set of fuzzy appraisal rules that returns a fuzzy emotion. Then, a defuzzification process returns the Pleasure and Arousal dimensions of the emotion that will be internally represented as a vector in a two-dimensional space. This vector is used to update the agent's mood according to the agent's personality. The agent can express this internal emotional state using a fuzzification process that translates the agent's mood into a fuzzy emotion. This fuzzification process uses the results of an experiment to generate a fuzzy emotion that is adapted to the cultural environment in which the agent is located. This methodology can be easily adapted to other languages.

4.1 Introduction

Over the last few years, the way in which humans interact with machines has been changing. The emergence of virtual assistants for both mobile phones and home is helping to facilitate how humans communicate with machines. However, human beings are emotional beings and these interactions can be improved by taking into account human emotions. Emotions affect the way in which we interact with others and establish links that will be the basis of future relationships. Thus, a computer system that can express emotions and understand the emotions of its human interlocutor will be able to improve its interactions with humans.

Affective computing [170] is the area of computation related to emotions that is based on theories of the sciences of psychology and cognition. One of the main goals of affective computing is to design computational models of affect to simulate human emotional behavior in a realistic way. Thus, different proposals have been designed from the computational [81, 148, 186] and psychological [201] perspectives. However, in general, most of these proposals are simplifications of psychological theories that were not originally proposed to be incorporated into computational models, thus reducing emotions to simple labels (as in emotion recognition) [103]. When this label-based representation is used, some inherent properties of emotions, such as the intensity or the proximity to other emotions, is lost. In this paper, we propose an internal representation of emotions using a continuous multidimensional space that is more appropriate for use in computational models since it provides a great capacity to represent emotions and mood, and we analyze their variations when receiving both internal and external stimuli.

Considering that humans use expressions like "very happy" or "a little sad", in recent years, some authors have proposed the use of fuzzy logic [238] to model affective processes since fuzzy logic is closer to the way in which human beings express their emotions [116]. On the other hand, another important difficulty when studying emotions is that the internal way of feeling emotions is highly dependent on the person's personality and the cultural environment. Different authors have shown that emotions depend on language and culture [196, 227]; specifically, the same emotion name/label can

be interpreted in a different way in different cultures. Moreover, different people show different amounts of spoken affect according to their language and cultural characteristics. That means that any appraisal process used in an affective agent should be adapted to the culture and language in which the agent will be used. Therefore, when computational models of affect are designed, the cultural environment in which they will be used must be considered. However, even though cultural and language factors are taken into account in other domains such as personal assistants, to our knowledge, there are still no affective computing models that really take cultural and language factors into consideration.

This paper presents a new appraisal model for a multi-agent affective BDI architecture [180] based on fuzzy logic that is adapted to cultural and language factors. Our proposal is composed of two independent processes. First, in the *Event Appraisal* process, an event is evaluated using fuzzy rules, producing a fuzzy emotion as a result. Second, the *Affect Adaptation* process defuzzificates the appraised emotion obtaining its *Pleasure* and *Arousal* dimensions, which are adapted to the cultural and language in which the emotion was elicited. These dimensions are used to update the agent's mood using the elicited emotion, taking into account the agent's personality and the language and culture in which the agent is located.

4.2   Previous work

Over the years, different theories have been developed to explain what emotions are and how emotions are elicited. Appraisal theories postulate that emotions are the result of an evaluation process that is triggered when an event occurs. From a theoretical perspective, the appraisal process is a response to a stimulus that triggers a series of processes that is in charge of evaluating that stimulus and, as a result, eliciting an emotion. In most cases, the appraisal process is explained using a set of variables known as *appraisal variables*. The number of emotions that can be generated in each appraisal theory depends on the number of appraisal variables used and the number of values that these variables can have. This way of eliciting emotions as the re-
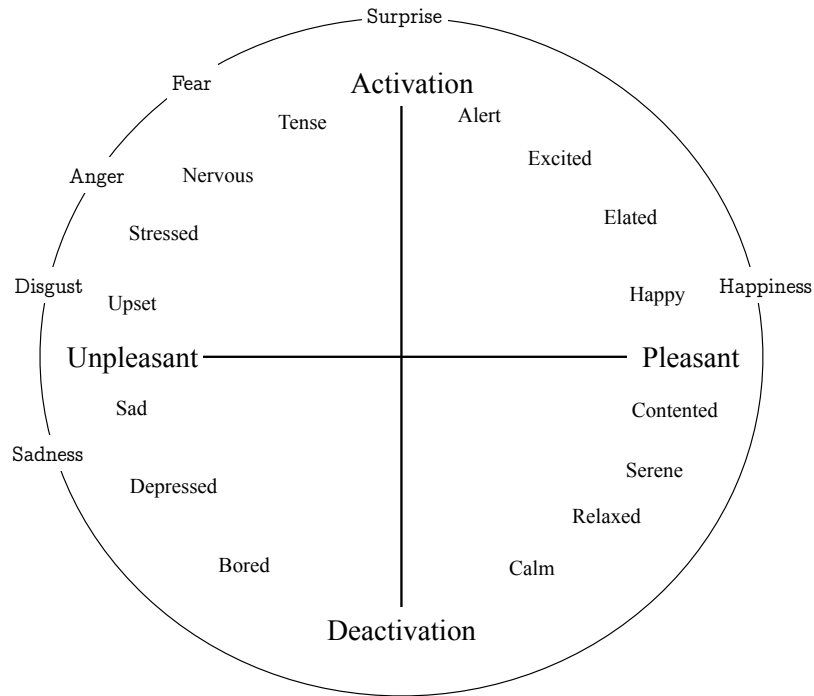
Figure 4.1: The circumplex model of affect: the outer ring contains the "prototypical emotion episodes" that match the basic emotions of Ekman's theory. Source: [195].

sult of the evaluation of a set of variables is very useful when using this model in a computer system. This is why, in recent years, most affective intelligent agent models have incorporated an appraisal process [6, 81, 148].

Different authors use different sets of appraisal variables to define the appraisal process. For example, Frijda [72] posits that the familiarity, expectedness, valence, controllability, agency, certainty, and importance of the detected event should be taken into consideration. The Scherer approach [201] makes a more detailed description of the factors involved in the appraisal processes such as predictability, urgency, power, or suddenness. Scherer also proposes a series of patterns that relate the appraisal variables to different emotions in order to propose an expert system for affective computing [200]. However, Scherer uses twenty-two appraisal variables, which makes this model

very complex to use. This is why this model has not been widely used in affective computing. In contrast, the OCC (Ortony, Clore, and Collin) model [163], which uses only eight appraisal variables, is the most commonly used model in affective computing. On the other hand, basic emotion theories, like the one proposed by Ekman [65], hold that there is a limited number of emotions (*Happiness*, *Surprise*, *Fear*, *Anger*, *Disgust*, and *Sadness*) and that each detected event elicits an associated emotion. According to this theory, these emotions are universally understood in different cultures and languages.

While Ekman argues that basic emotions are universal and transcultural, constructivist theorists, like Russell, argue that emotions depend on language and culture [196]. Russell's theory [195] is based on the fact that there is not always a direct correspondence between the words representing emotions in two different languages. For example, in the German language, there is an emotion called "*Sehnsucht*" [107], whose meaning is "a strong desire for an alternative life". In languages such as English or Spanish, there is no single word to express this emotion. Russell's theory is more consistent with findings in cognitive neuroscience [174]. These claim that emotions are a combination of a small number of dimensions rather than being directly related to events, as in Ekman's theory and other basic emotion theories [65], which affirm that emotions are related to an independent neuronal system. Russell relates emotions with a pair of values, *Pleasure* and *Arousal*, providing a two-dimensional space for the representation of emotions. In addition, through experimentation, he observed that emotions follow a circular pattern in these two dimensions, which led him to propose his best-known theory: *The Circumplex Model Of Affect* [193] (Fig. 4.1). Mehrabian proposed the PAD (*Pleasure*, *Arousal*, *Dominance*) model [154], which adds the *Dominance* dimension to represent emotions. This third dimension can help to clarify negative emotions. For example, the emotion of *Fear* is associated with a low level of *Dominance*, while the emotion of *Anger* is associated with a high level of *Dominance*. However, the *Dominance* dimension is not included within many models, and it is used as an appraisal variable instead [147]. Moreover, this dimension does not seem to have a high variability among emotions [107], and some authors have found that human beings have difficulty and show confusion when they try to assign a *Dominance* value to the emotions that they are feeling [122].

Other authors, such as Reisenzein [183], showed the relationship of the intensity of emotions with the variables of *Pleasure* and *Arousal*. In an experiment, Reisenzein observed that the levels of *Pleasure* and *Arousal* are related to the intensity of emotions. Apparently, the intensity of emotions is proportional to the levels of *Pleasure* and *Arousal* so that the higher these levels, the greater the intensity. For example, the intensity of the *Happiness* emotion is highly dependent on the level of *Pleasure*, while the intensity of the *Alert* emotion is more dependent on the level of *Arousal*. He also observed that certain minimum levels of *Pleasure* and *Arousal* are required in order for emotions to show themselves. Therefore, when both variables have low values, the absence of emotions can be considered.

The relation between emotions, personality, and mood

When modeling the appraisal process, the relation between emotions and other affective characteristics such as personality or mood must be taken into consideration. Although the difference between emotions and mood continues to be a topic of debate [19], in general, it is accepted that an emotion is a rapid response to a given stimulus, while mood has a longer duration (from minutes to days) and a lower intensity than emotions and is not related to any particular stimulus. Instead, mood is produced by a succession of stimuli and other factors such as the context or the person's personality [64]. Empirical evidence suggests that there is a critical impact of emotions on cognition and a high variability of this impact among individuals with different personality factors. Personality can make a person more or less likely to experience certain types of mood [133].

One of the most commonly used models to define personality is the five-factor model (FFM) [152] in which personality is defined using five variables (or traits): *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*. Different works that are based on the FFM relate personality factors with cognitive abilities, behaviors, and social skills. For example, there is evidence [56] that the *Extraversion* factor predisposes one to experience a more positive affect more frequently and with greater intensity, appraising emotions such as joy, enthusiasm, or energy. In contrast, *Neuroticism* pre-

disposes one to experience a negative affect as well as suffering from negative moods such as fear, sadness, depression, or anger.

Emotions in affective agents

In recent years, different models have been proposed to simulate emotions using intelligent agents [5, 47, 186]. The proposed model for the EMA affective agent [148] is based on the subsequent formalization of the theory presented by Smith and Lazarus [135] that was carried out by Gratch and Marsella [88]. They propose a one-level appraisal process that takes into account past, current, and future events. They define different thresholds for the appraisal variables to elicit emotions [88]. In this model, intensity is defined as the product of different numeric appraisal variables. For example, the product of desirability and likelihood is used to estimate the intensity of the *Joy* emotion. The appraisal process of EMA has been used as a base to define emotion elicitation processes in other affective agent models such as *GenIA³* [6].

Another interesting approach is the one proposed by Gebhard in ALMA (A Layered Model of Affect) [81]. This model is based on the OCC appraisal theory [163] and models appraisal using subjective appraisal rules; therefore, each agent can have different appraisal rules. In these rules, the different appraisal variables are bounded by numeric values for eliciting each emotion. The authors also propose a decay rate for emotions, which decreases the intensity of the emotion over time until the emotion disappears.

Generally, humans express emotions and moods through terms such as "very happy" or "a little happy". Therefore, in the last few years, some authors have proposed using fuzzy logic to define affective computational models. For example, the appraisal model for FLAME [67], which is based on the Roseman [190] and OCC theories, uses variables with fuzzy values. In this approach, each emotion is selected based on a series of simple rules. Another interesting approach is the one presented by Jain et al. in EMIA [108]. The authors propose a fuzzy logic emotional model for five of Ekman's basic emotions (all except the *Disgust* emotion) [65]. The EMIA model implements a simplification of the Scherer model using five appraisal variables.

The appraisal process is affected by the agent's memory and allows more than one emotion to be elicited in each appraisal cycle. All of the emotions have a fuzzy intensity value, and this intensity is modified over time by an emotion decay process.

Other affective characteristics, such as personality or mood, have also been used in affective computing to define the emotion appraisal process. Personality is usually used to create behaviors in multi-agent systems [39, 215]. For example, in [38], the authors propose that personality and mood have an effect on the appraisal variables and thus on the elicited emotions. This model proposes that personality affects the ranges of the fuzzy appraisal variables. Similarly, in [39], a model that uses fuzzy appraisal variables is proposed based on the *Integrative Model* [189]. That model is designed to adapt the cognitive and affective capacities of an agent to internal and external factors among which gender or personality are considered. To this end, each appraisal variable is defined with a membership function that is adapted according to these internal and external factors. The modification of the membership function of each appraisal variable is done both by modifying the type of function (e.g., pyramidal, trapezoidal) and by modifying the parameters of the membership function. However, at the moment, this model is only a proof of concept.

Discussion

When affective agents are modeled to simulate emotional behavior, the most important task is to choose the right emotion when an event is triggered. This is because the generated emotion will affect the cognitive processes of the agent, such as reasoning or the selection of behavior. The selected emotion will also have an impact on affective characteristics such as mood or empathy. Therefore, generating an incorrect emotion when an event occurs can cause erroneous behaviors that decrease the realism of the emotional behavior simulation. To model this process, several variables must be taken into account, such as the agent's concerns or the desirability of the event and other factors such as personality or mood. Several affective models use numerical variables, but this is problematic because it is somewhat different from how

human beings express their emotions; humans find it very difficult to quantify
emotions with numerical values.

In our model, we have decided to use fuzzy logic as a more natural way
of eliciting emotions in the appraisal process. Models that use fuzzy logic are
more suitable for simulating the human appraisal process due to the "fuzzy"
way of expressing emotions by human beings. The use of fuzzy logic helps the
understanding of the emotional states of the human beings that are interact-
ing with the agents. In recent years, some proposals for emotion elicitation
processes have been defined using fuzzy logic. For example, FLAME [67] pro-
poses a small set of rules but involves the user too much in its implementation.
In addition, even though the FLAME model allows emotions with intensity
to be generated, the simplicity of the rules makes it a very basic model. For
example, the *Joy* emotion is elicited if a desirable event occurs without taking
into account the rest of the appraisal variables. Therefore, the rules are too
simple, assigning the selected emotion by using only one appraisal variable.
In addition, the appraisal variables are evaluated in a binary way: desirable or
undesirable. Therefore, there are only a few combinations of values available
to generate different emotions. Thus, this model has a very limited catalog of
emotions, which, as stated above, hinders the user experience. Other fuzzy
models, such as the model presented in EMIA [108], proposes a fuzzy logic
approach for the appraisal process. That model reduces Scherer's model to
five appraisal variables. However, the final model has two hundred rules to
define an appraisal process that generates only five emotions. Therefore, even
though the FLAME model is very basic due to the small number of rules, the
EMIA model is too complex given its enormous number of rules. In addi-
tion, the EMIA model does not consider mood or personality in the appraisal
process.

When simulating human affective behavior, the effect that different exter-
nal and internal factors (e.g., personality or mood) have on cognitive pro-
cesses must be taken into account. These factors have a direct consequence
when evaluating an event, biasing the appraisal processes and consequently
influencing emotions or mood. Some current proposals, such as the one pre-
sented in [38], use personality and mood to influence emotion selection in the
appraisal process. However, in this model, the intensity of emotion is not

contemplated. Therefore, this model becomes much simpler to design and implement. This can have a negative effect on the behavior of the agent since the catalog of emotional responses decreases. Other proposals such as the one proposed in [39] take into account the importance of these internal and external factors, but that model has only been tested with the gender factor. Furthermore, although it is accepted that factors such as personality have an effect on emotions, which predisposes the individual to certain types of emotion, there is not enough evidence to determine what the effect of personality is on an appraisal variable.

As we have stated above, there is a large number of appraisal theories that try to explain how emotions are generated. Each theory has a series of advantages and disadvantages. For example, the patterns that Scherer [200] proposes in his computing model of affect are interesting when determining which variables affect each emotion. However, Scherer uses a large number of variables, which makes this model very complex for development in an affective agent. Furthermore, in this model, the intensity of the emotions is not considered. This may affect the behavior of the agent since the catalog of actions of the agent will be very limited because the intensity of the emotion cannot be taken into account to decide the action to be executed. A small catalog of emotions can affect the interactions with human users because, if an agent shows the same emotions very often, the agent loses credibility and the users tend to become bored. Therefore, the more we expand the catalog of emotions of an agent, the greater the number of different emotional behaviors that can be simulated.

The internal representation of emotions must also be taken into account when defining an emotional model for an affective agent. Most of the emotional models proposed to date reduce the representation of emotions to simple labels [69]. However, dimensional representations seem to be more appropriate for developing computational models since they allow emotions to be stored using variables instead of simple labels and variables can hold a wider range of values. The Circumplex Model of Affect [193] (Fig. 4.1) has already been used in affective computing [198]. Nevertheless, the schema proposed by Russell is only a reinforcement for his theory. In this scheme, he uses a simplified spacial location for each emotion, but this is unnatural, and, conse-

quently, an agent using this simplified representation of emotions could show erratic behavior that is very different from the real behavior of a human being. Therefore, performing a direct use of this model in a computational system could lead to unreal emotional behaviors.

In the next section, we introduce our new fuzzy appraisal model based on representing emotions in a *Pleasure-Arousal* space. Section 4.3 describes how a fuzzy rule-based system selects one or more emotions by evaluating the appraisal variables. Section 4.3 presents the *Pleasure-Arousal* space where agents can internally represent the appraised emotions using a defuzzification function. This internal representation of emotions allows the agent to update its mood as presented in Section 4.3. Finally, when the agent needs to communicate or express its internal emotional state, a fuzzification process (presented in Section 4.3) translates the internal emotional state represented in the *Pleasure-Arousal* space into a fuzzy emotion. The fuzzification process uses the results of an experiment to generate a fuzzy emotion that is adapted to the language and cultural environment of the human interlocutor of the agent. The article ends with conclusions and future works.

4.3 A model for emotion elicitation in affective agents

The appraisal process is one of the most important processes in an affective agent because suitable agent behavior depends on the emotion selected. A good definition of this process is even more important when the agent is designed to interact with humans since erratic or erroneous behavior of the agent can make the human become frustrated. In addition, processes such as the selection of the action to be executed or the updating of the agent's mood will depend on the emotion elicited.

As we have discussed above, there are different psychological theories to explain emotional processes. Appraisal theories provide an explanation of how an emotion is generated when an event is perceived. Alternatively, other theories, such as Russell's theory [193], offer the possibility of relating emotions with dimensional variables. We propose a fuzzy model of emotions that

adapts the advantages of the appraisal models for selecting an emotion when an event occurs and the advantages of representing the emotion in a multidimensional space. As a novelty, by using this multidimensional representation, our model can adapt emotions to the agent's affective characteristics and to the culture and the language in which the agent will be located.

Our model is composed of two processes. First, in the *Event Appraisal* process, an event is evaluated by our fuzzy appraisal algorithm and produces an emotion which we will refer to as *Appraised Emotion*. This fuzzy process allows us to generate emotions and their intensity, capturing the kind of uncertainty that humans beings use when expressing their emotions. Second, in the *Affect Adaptation* process, the mood is updated based on the *Appraised Emotion*, the agent's personality, and the culture and language in which the agent is located. The following section describes both processes in detail.

The Event Appraisal process

The *Event Appraisal* process defines a fuzzy appraisal process to select an emotion when the agent perceives an event. The selected emotion corresponds with one of the six basic emotions proposed by Ekman [65]: *happiness*, *disgust*, *sadness*, *anger*, *fear*, and *surprise*. Moreover, based on the emotions proposed by Russell in [196], two emotions, *calm* and *boredom*, have also been incorporated to represent the lack of interactions between the agent and the user and also the lack of relevance of the event.

We have designed a fuzzy rule-based system [122] that, depending on the value of the appraisal variables, returns the elicited emotion and its intensity. The fuzzy values used to define the rules are based on the computational model proposed by Scherer [200], but we solve the main weakness of this model: Our model needs a smaller number of appraisal variables and it also calculates the intensity of the elicited emotion. Moreover, in our model, the values of each appraisal variable are associated with the intensity of the emotions as proposed by OCC[1]. According to this theory, two of the most relevant appraisal variables attributed to emotions produced by events, such as hap-

---

[1]The relationship between the appraisal variables of both the OCC and the Scherer model can be found in [200]

piness or sadness, are desirability and likelihood. Therefore, in our model, we associate these two variables with Scherer's model to obtain the label and the intensity of the emotions using fuzzy variables. The appraisal variables selected for our model are the following:

- **Expectedness**, which represents the agent's level of expectation associated with the event.

- **Likelihood**, which represents the estimated probability of occurrence of the event.

- **Desirability**, which represents how desirable or undesirable the event is for the agent; it can be positive or negative.

- **Causal attribution**, which represents who is responsible for the event.

- **Controllability**, which represents the number of possible plans that the agent has to deal with for that event.

The use of fuzzy logic in the appraisal process facilitates the definition of fuzzy rules to generate emotions since fuzzy logic is closer to the way in which human beings express their emotions. Therefore, our fuzzy logic-based approach produces better and more reliable results than other non-fuzzy appraisal processes.

In our appraisal model, each event is associated with a tuple:

$$(4.1) \qquad\qquad event \leftarrow \langle e, l, d, a, c \rangle$$

where e, l, d, a, and c represent the fuzzy values for the appraisal variables *Expectedness*, *Likelihood*, *Desirability*, *Causal_attribution*, and *Controllability*, respectively. The possible values for these variables are:

$$(4.2) \quad \begin{aligned} &e, l, c \in \{Low, Medium, High\} \\ &d \in \{High\_desirable, Desirable, Low\_desirable, Low\_undesirable, \\ &\qquad Undesirable, High\_undesirable\} \\ &a \in \{Self, Other\} \end{aligned}$$

On the other hand, an emotion is defined by a tuple:

$$(4.3) \qquad emotion = \langle type, int \rangle$$

where *type* is the label representing the emotion type, and *int* represents
the fuzzy intensity of the emotion:

$$(4.4) \quad \begin{aligned} type \in \{ & Happiness, Disgust, Sadness, Anger, Fear, Surprise, \\ & Calm, Boredom \} \\ int \in \{ & Strong, Medium, Light, Neutral \} \end{aligned}$$

The *Neutral* value is the default value for emotions that have not been
elicited in the appraisal process. In addition, we have introduced a new
appraisal variable named *Time Without Events* (TWE). This variable is not
part of the event; it is an internal variable of the agent that measures the
time that has passed since the last event. We have proposed this appraisal
variable to elicit the *calm* and *boredom* emotions. The possible values for this
variable, represented by t, are:

$$(4.5) \qquad t \in \{ Low, Medium, High \}$$

We use a fuzzy rule-based system to calculate the elicited emotion and its
intensity. The general structure of a fuzzy rule [158] is:

(4.6)

r_i : IF e is x_e^i and l is x_l^i and d is x_d^i and a is x_a^i and c is x_c^i and t is x_t^i

THEN *type* is y_{type}^i and *int* is y_{int}^i

where r_i is the ith fuzzy production rule, x_e^i represents the fuzzy value
for the *Expectedness* appraisal variable in the ith rule, x_l^i is the fuzzy value
for *Likelihood*, x_d^i is the fuzzy value for *Desirability*, x_a^i is the fuzzy value for
Causal attribution, x_c^i is the fuzzy value for *Controllability*, x_t^i is the ith fuzzy
value for *Time Without Events*, and y_{type}^i and y_{int}^i are the output values for
the emotion type and the intensity calculated by the ith rule (see Table 4.1).

For example, a fuzzy rule appraising the Fear emotion with High intensity
is defined as:

(4.7)
r_{13} : IF e is Low and l is High and d is High_Undesirable and a is Other
and c is Low and t is Low
THEN $type$ is Fear and int is High

Table 4.1 summarizes the parameters of the 24 fuzzy rules that are pro-
posed to define the appraisal process. Each row corresponds to a fuzzy rule,
while each column contains the possible fuzzy values for each appraisal vari-
able that is associated with each emotion. As we stated above, the definition
of the appraisal variable values is based on the model proposed by Scherer
[200] and the OCC model [163].

We have defined three fuzzy values for the intensity of each emotion. This
intensity depends on the fuzzy values of the set of appraisal variables [148].
This fuzzy process not only derives emotions and their intensity from the set
of appraisal variables, it also unifies the selection of the type and the intensity
of the emotion in the same rule. Therefore, it is not necessary to define a new
process to derive the intensity as in other models [67, 148].

The Affect Adaptation process

Two important factors that must be taken into consideration when modeling
affective agents are mood and personality. Mood can be seen as the result
of the various emotional events that are produced during a period of time.
On the other hand, personality influences the type and intensity of the mood
of a person [133]. For example, a person with high levels of neuroticism will
be more likely to experience negative moods more frequently and with more
intensity. In our model for affective agents, mood is represented in a two-
dimensional space in terms of *Pleasure* and *Arousal*. The *Affect Adaptation*
process updates the mood by modifying the values of *Pleasure* and *Arousal*
according to the current emotion, the agent's personality, and the language

Table 4.1: Fuzzy model for estimating the emotion and its intensity. Where HD = High desirable, D = Desirable, LD = Low Desirable, LU = Low Undesirable, U = Undesirable, and HU = High Undesirable.

Emotion	Intensity	Expectedness	Likelihood	Desirability	Causal attribution	Controllability	TWE
Happiness	High	Medium	High	HD			
	Medium	Medium/Low	High/Medium	D			
	Low	Medium/Low	High/Medium	LD			
Disgust	High	Low	High	HU	Other	High/Medium	
	Medium	Low	High	U	Other	High/Medium	
	Low	Low	High	LU	Other	High/Medium	
Sadness	High		High	HU/U		Low	
	Medium		High/Medium	U		Low	
	Low		High/Medium/low	LU		Low	
Anger	High	Medium	High	HU		Medium	
	Medium	Medium	High/Medium	U		Medium	
	Low	Medium	High/Medium/Low	LU		Medium	
Fear	High	Low	High	HU	Other	Low	
	Medium	Low	High/Medium	U	Other	Low	
	Low	Low	High/Medium/Low	LU	Other	Low	
Surprise	High	Low	Low		Other		
	Medium	Low	Medium		Other		
	Low	Low	High		Other		
Calm	High					Medium	Low
	Medium					Medium/Low	Medium
	Low					Medium/Low	High
Boredom	High					Low	Low
	Medium					Low	Medium
	Low					Low	High

and culture in which the agent is located. We describe the different phases of this process in the following sections.

Representing emotions in a Pleasure-Arousal space

As we have shown in the previous sections, the result of the *Event Appraisal* process is the *Appraised Emotion*, which is composed of an emotion type and a fuzzy intensity. In the *Affect Adaptation* process, we use the *Pleasure*

and *Arousal* dimensions of the *Appraised Emotion* to adapt the agent's mood. Therefore, to calculate how the mood will be modified by the *Appraised Emotion*, we need to represent this emotion in the same two-dimensional space that is used for representing the agent's mood. To avoid the loss of information, our multidimensional model must be able to represent both the emotion type and the emotion intensity.

Our proposal represents a fuzzy emotion in a two-dimensional space that is adapted to the culture and language in which the agent will be used. This model is based on the results of the experiment presented in [219]. In this experiment, one-hundred European Spanish-speakers, 40 females and 60 males ranging in age between 18 and 60 years old, were asked to assign fuzzy values of *Pleasure* and *Arousal* to ten terms/words expressing emotions in their mother tongue. The emotions selected for this experiment were: *Happiness*, *Excitement*, *Surprise*, *Fear*, *Disgust*, *Anger*, *Sadness*, *Boredom*, *Calm*, and *Sleepiness*. When the *Pleasure* and *Arousal* dimensions have low levels, the intensity of the emotion is so low that it can be considered that there is no emotion. Therefore, we discarded those responses whose values of *Pleasure* and *Arousal* were close to zero. We obtained the mean of the *Pleasure* and *Arousal* for each emotion, and then we calculated the mean and the standard deviation of the angle of each emotion in this two-dimensional space. This angle represents the meaning that each participant associates to each emotion. The results of the experiment are summarized in Table 4.2.

Our model also divides the *Pleasure-Arousal* space into four degrees of intensity: *Strong*, *Medium*, *Light*, and *Neutral* (Fig. 4.2). We have added the fuzzy value *Neutral* to refer to those emotions that are not intense enough to be considered as elicited. When we superimpose the results of the experiment on the model proposed for the intensity of emotions, we obtain different areas that relate the *Pleasure* and *Arousal* dimensions to each emotion label and its intensity adapted to the Spanish language. Fig. 4.3 shows the emotion model proposed in this paper, where the area assigned to each emotion, using one standard deviation, is represented in a different color. This model can be easily adapted to other languages and cultural environments using the results of similar experiments to assign different areas to each emotion. An agent with more than one *Pleasure-Arousal* space can easily adapt its emotion
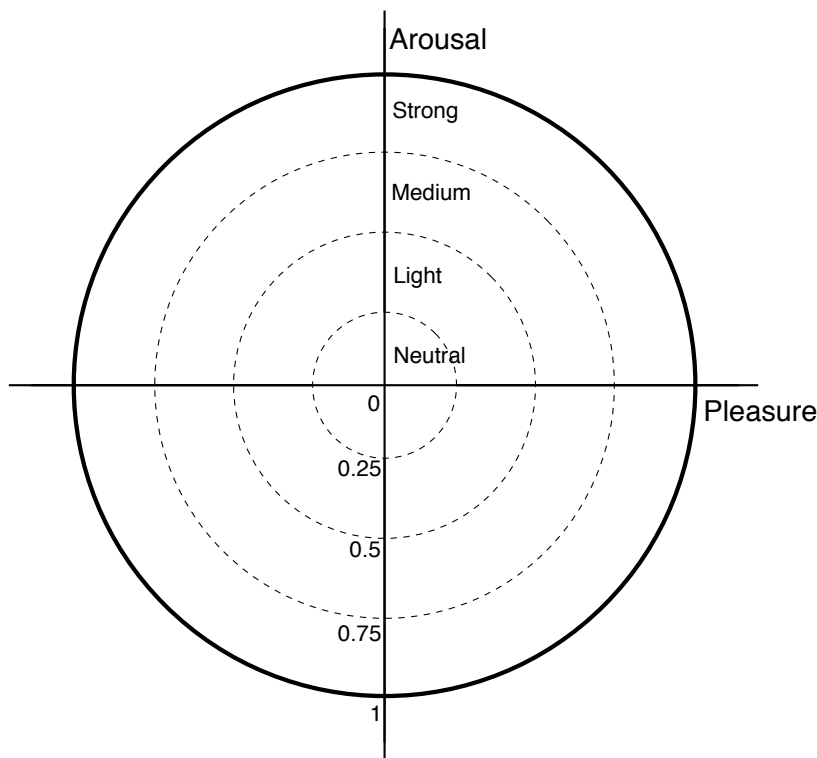
Figure 4.2: The proposed fuzzy model for the intensity of the emotions in the *Pleasure-Arousal* space.

expression depending on the language and culture of its interlocutor without modifying the rest of its affective components.

We have designed a defuzzification function that translates the fuzzy appraised emotion into its *Pleasure* and *Arousal* equivalent values. The defuzzification function determines the direction of the *Appraised Emotion* vector $\vec{e}$ in this culturally located, two-dimensional space using the mean angles obtained in the experiment for each emotion (see Table 4.2). For example, if the *Appraised Emotion* is *Anger*, the direction of the corresponding vector $\vec{e}$ will be 138.5 degrees, which corresponds to the mean angle obtained in the experiment for the *Anger* emotion . The defuzzification function also calculates the vector modulus, which indicates the intensity of the emotion, as the mean value for each intensity. For example, if the intensity of the *Appraised Emotion* is *Strong*, the modulus of the emotion vector $\vec{e}$ will be 0.875, which

Table 4.2: Results of the experiment expressed in degrees.

Emotion	Pleasure	Arousal	Mean angle	Standard deviation
Happiness	0.90	0.42	25.09	19.02
Excitement	0.76	0.64	39.97	10.32
Surprise	0.31	0.95	71.63	26.38
Fear	-0.58	0.81	125.51	15.61
Anger	-0.74	0.66	138.55	16.90
Disgust	-0.99	-0.04	182.58	43.65
Sadness	-0.96	-0.27	196.02	22.48
Boredom	-0.41	-0.91	245.34	21.41
Sleepiness	-0.11	-0.99	263.59	15.56
Calm	0.74	-0.67	318.12	35.89

corresponds to the midpoint of the *Strong* intensity that goes from 0.75 to
0.1. The result of this defuzzification function is the vector $\vec{e}$, which represents the *Appraised Emotion*. Considering that the horizontal axis in this
two-dimensional space represents *Pleasure* and the vertical axis represents
Arousal, if α is the angle of the vector $\vec{e}$, then the *Pleasure* and *Arousal*
values of the *Appraised Emotion* can be calculated by a simple trigonometry
formula as follows:

$$(4.8) \qquad (Pleasure, Arousal) = (|\vec{e}| \cdot \cos\alpha, |\vec{e}| \cdot \sin\alpha)$$

For instance, let us suppose an agent a receives an event. The fuzzy
appraisal process evaluates the values of the appraisal variables associated
to that event and obtains the emotion e defined by the type *Anger* and the
intensity *Strong* by using the fuzzy rule-based system. The direction of the
vector $\vec{e}$ corresponding to the *Anger* emotion is $\alpha = 138.5$ (as shown in
Table 4.2), and its modulus is $|\vec{e}| = 0.875$. The agent can calculate the
Pleasure and *Arousal* values associated with the emotion vector $\vec{e}$ using
Formula 4.8: $(Pleasure, Arousal) = (-0.66, 0.58)$.

We have shown how our model internally represents an emotion by a vector
in which the modulus represents the intensity of the emotion and the direction

represents the type of emotion. The result is a circular representation of the emotions with some similarities with the scheme proposed by Russell (Fig. 4.1). However, as we stated above, Russell's scheme should not be used directly since it is only a scheme that was used to reinforce the idea that emotions follow a circular pattern; the area assigned to each emotion is more restricted, and the emotion intensity is not considered.

Updating the mood

Once the agent obtains the *Pleasure* and *Arousal* values of the appraised emotion, the agent can use these values to update its mood according to the agent's personality. In this phase, the vector of the agent's mood, which is also represented in the *Pleasure-Arousal* space, is "attracted" by the emotion vector to a greater or lesser extent depending on the agent's personality. We have defined two functions, called *Mood_Resistance* and *Emotion_Influence*, to model this behaviour. The *Mood_Resistance* function determines how a mood m resists being changed by the appraised emotion. The *Emotion_Influence* function determines the influence or force of the emotion e to modify the agent's mood. Both functions depend on the agent's personality. For example, an agent with a personality that is prone to negative affect will have a *Mood_Resistance* for the negative moods that is higher than the *Mood_Resistance* for the positive moods. That means that the agent will be more prone to suffer a negative mood with greater intensity than an agent with a personality that is prone to positive affect [216].

The current mood of an agent at instant t, $\overrightarrow{m}_t$, will be updated as a combination of the previous mood vector (mood at instant $t-1$), $\overrightarrow{m}_{t-1}$, weighted by the *Mood_Resistance(m)* and the emotion vector $\overrightarrow{e}$ weighted by the *Emotion_Influence(e)* function:

$$(4.9) \quad \overrightarrow{m}_t = \overrightarrow{m}_{t-1} \cdot Mood\_Resistance(m_{t-1}) + \overrightarrow{e} \cdot Emotion\_Influence(e)$$

where *Mood_Resistance(m_{t-1})* and *Emotion_Influence(e)* represent how the personality of the agent affects its mood. These two functions are calcu-

lated using the following formulas:

$$(4.10) \qquad Mood\_Resistance(m) = \frac{\sum_{p \in P} \beta_p \cdot \theta_{m,p}}{\sum_{p \in P} \theta_{m,p}}$$

$$(4.11) \qquad Emotion\_Influence(e) = \frac{\sum_{p \in P} \beta_p \cdot \theta_{e,p}}{\sum_{p \in P} \theta_{e,p}}$$

where P represents the set of personality traits of an agent. For example, for the five-factor model (FFM), these traits are *Openness* (O), *Conscientiousness* (C), *Extraversion* (E), *Agreeableness* (A), and *Neuroticism* (N); therefore, P is defined by the set $P = \langle O, C, E, A, N \rangle$. β_p is the value of a personality trait $p \in P$. $\theta_{e,p}$ is a weight that relates the personality trait p with the emotion e. There will be one weight $\theta_{e,p}$ for each pair of emotion type and personality trait (e, p), but many of these weights could be zero. Each weight $\theta_{e,p}$ determines how the personality trait p influences the emotion type e. Therefore, the set of weights $\theta_{e,p}$ can be viewed as the set of correlations between emotions and personality traits: the greater the correlation between the personality trait p and the emotion type e, the greater the value of $\theta_{e,p}$. For example, considering that the trait of *Extraversion* E is related to positive emotions and *Neuroticism* N does not have a very high relation with emotions of this type, for the *Happiness* emotion, the value of the weight $\theta_{Happiness,E}$ will be greater than the value of the weight $\theta_{Happiness,N}$. Different $\theta_{e,p}$ values can be chosen by the agent programmer to generate agents with different emotional behaviors and personalities.

To better understand this method, let us reconsider the previous example where after evaluating an event at instant $t - 1$, an agent $a1$, has elicited the *Appraised Emotion* represented by vector $\vec{e}$ in Fig. 4.3. Let us consider that the mood for agent $a1$ at instant $t - 1$ is defined by the vector $\overrightarrow{m}_{t-1,a1}$. The *Pleasure* and *Arousal* components for the mood vector $\overrightarrow{m}_{t-1,a1}$ are $(-0.30, -0.20)$. Let us also consider a second agent, $a2$, who has appraised the same emotion as agent $a1$ ($\vec{e}$) and has the same mood. To simplify, we will represent the mood of both agents by $\overrightarrow{m}_{t-1}$ ($\overrightarrow{m}_{t-1} = \overrightarrow{m}_{t-1,a1} = \overrightarrow{m}_{t-1,a2}$). The

Table 4.3: Example of $\theta_{e,p}$ values.

Emotion	E	N
Anger	$\theta_{Anger,E} = 0.5$	$\theta_{Anger,N} = 0.8$
Sadness	$\theta_{Sadness,E} = 0.6$	$\theta_{Sadness,N} = 0.7$

Affect Adaptation process will update the agents mood at instant t taking into account the effect of the *Appraised Emotion*, the mood of the agents at instant $t-1$, and the personality of the agents. To simplify this example, the agent's personality is defined using only two traits: *Extraversion* (E), which is related to positive affect; and *Neuroticism* (N), which is related to negative affect. Therefore, the personality in this example is defined by the tuple:

$$(4.12) \qquad\qquad P = \langle E, N \rangle$$

On the other hand, for this example, we consider the values for $\theta_{e,p}$ shown in Table 4.3. These values are based on the results obtained from the experiments presented in [75], making them consistent with theories that associate *Extraversion* with positive emotions and *Neuroticism* with negative emotions [206]. Taking into account these values, if we define the personality of agent $a1$ by a level of *Extraversion* of 0.9 and a level of *Neuroticism* of 0.1 ($P_{a1} = (0.9, 0.1)$) and the personality of agent $a2$ by a level of *Extraversion* of 0.1 and a level of *Neuroticism* of 0.9 ($P_{a2} = (0.1, 0.9)$), then the values of *Pleasure* and *Arousal* for the new mood of agents $a1$ and $a2$ at instant t calculated by Formula 4.9 are $\overrightarrow{m}_{t,a1} = (-0.4, 0.1)$ and $\overrightarrow{m}_{t,a2} = (-0.6, 0.3)$. These values are represented in Fig. 4.3 by the vectors $\overrightarrow{m}_{t,a1}$ and $\overrightarrow{m}_{t,a2}$.

Comparing the resulting moods of both agents, it can be observed that, in the case of agent $a1$, which has a high level of extraversion, the mood has been less affected by a negative emotion than in the case of agent $a2$, which has a higher level of neuroticism. These results are consistent with theories that claim that personalities with low *Extraversion* and high *Neuroticism* are more prone to negative moods [27, 56].

As we have shown in our proposal, the mood is adapted to the current emotion according to the agent's personality. Therefore, our model facilitates the development of emotional multi-agent systems since it allows different moods

to be easily obtained for each agent based on the agent's personality. Different moods will produce different agent behaviors, which will improve the agent's capability to simulate human behaviors [216]. For example, an agent with a negative personality will be more prone to negative moods than an agent with a positive personality. Therefore, from psychological and neurological perspectives, this proposal is more consistent with the different theories that argue that emotions and personality have an effect on mood [133] and it also explains why people are more or less likely to have certain types of mood.

The Fuzzification Process

In our model, the pair $(Pleasure, Arousal)$ is internally used by an agent to represent the appraised emotion and its mood in the two-dimensional space. When the agent needs to express its mood, it has to perform a fuzzification process to determine the emotion type and intensity corresponding to the mood vector defined by the pair $(Pleasure, Arousal)$. The intensity of the emotion is easily calculated using the modulus of the emotion vector and the fuzzy model represented in Fig. 4.2. On the other hand, the process to calculate the emotion type corresponding to an emotion vector can be viewed as a classification problem where there is a set C of ten classes that correspond to the ten emotions.

$$
\begin{aligned}
C \in \{ & Happiness, Excitement, Surprise, Fear, Anger, Disgust, \\
& Sadness, Boredom, Sleepiness, Calm \}
\end{aligned}
$$
(4.13)

Using the Gaussian models obtained from the experiment for Spanish-speakers (Table 4.2), we can estimate the probability of each emotion. We have defined the $t(\overrightarrow{v})$ function as a Bayesian classifier that returns the emotion type with the maximum likelihood for an angle α of the emotion vector $\overrightarrow{v}$:

$$
t(\overrightarrow{v}) = \arg \max_{c} \widehat{P}(C = c \mid \alpha)
$$
(4.14)

where $c \in C$, and $\widehat{P}(C = c \mid \alpha)$ represents the estimated conditional probability of the class of emotion c for the angle α. This probability is
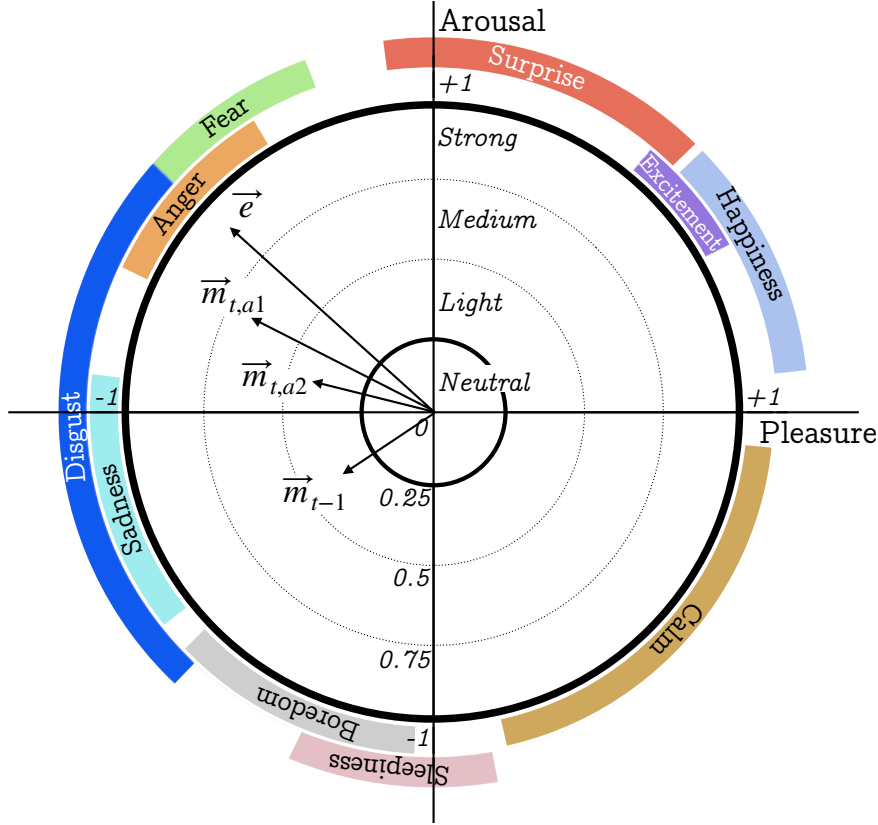
Figure 4.3: Example of the estimation of the mood considering two different personalities.

estimated by the normal distributions obtained in the experiment as:

$$\widehat{P}(C = c \mid \alpha) \sim f(\alpha \mid \mu_c, \sigma_c) \tag{4.15}$$

where μ_c and σ_c represent the mean and the standard deviation of the class of emotion c. $\widehat{P}(C = c \mid \alpha)$ is estimated using the probability density function of a Gaussian distribution $f(\alpha \mid \mu_c, \sigma_c)$ [112]:

$$f(\alpha \mid \mu_c, \sigma_c) = \frac{1}{\sigma_c \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\alpha - \mu_c)^2}{2 \cdot \sigma_c^2}} \tag{4.16}$$

Continuing with the previous example, if agent a wants to express the emotion $\vec{e} = (-0.66, 0, 58)$ (see Figure 4.3), the agent will calculate the most probable labels for the type and the intensity of the emotion through the

Fuzzification Process. First, the modulus of the vector is calculated (in this example, $|\vec{e}| = 0.875$) to obtain the fuzzy intensity. In this case, since the module is in the range of the *Strong* intensity ($0.75 \leq |\vec{e}| \leq 1$), the label of the emotion intensity is *Strong*. Then, the label of the emotion type is obtained using Formula 4.14 and the Gaussian models obtained from the experiment in the same cultural environment in which the agent is located. In this case, the maximum argument for the function $t(\vec{e})$ is $\widehat{P}(C = Anger \mid \alpha = 138.5)$. Therefore, the most likely label for that emotion is *Anger*.

Traditionally, in affective computing, most of the proposed appraisal models calculate the intensity and the emotion type of emotions independently by using numerical variables. Our model unifies the appraisal of the type and intensity of the emotion in the same rule. Dimensional models, such as the one proposed in this paper, also allow the affective behavior of the agent to be described in a more precise way. This dimensional representation defines regions where emotions can occur and therefore represents the proximity between two emotions and the intensity of the emotions. This representation allows the simulation of processes such as the decay rate of the mood over time and opens the doors to the simulation of new processes such as empathy or emotional contagion.

4.4   Conclusions and future work

In this paper, we have proposed a new fuzzy appraisal model using the *Pleasure* and *Arousal* dimensions, which is adapted by a defuzzification process to the culture and language in which the agent is located. Our model uses fuzzy logic to better reproduce the way in which humans express emotions. We propose two processes to simulate affective behavior. When an event is perceived, the *Event Appraisal* process evaluates five fuzzy appraisal variables to obtain the fuzzy *Appraised Emotion*. Then, the *Affect Adaptation* process defuzzificates the *Appraised Emotion* and adapts the mood using this defuzzificated *Appraised Emotion*, the agent's personality, and the cultural environment. This *Affect Adaptation* process allows different agents to have different moods when they perceive the same event, which will allow agents

to show different behaviors depending on the personality of each agent.

The proposed appraisal model has two important advantages. It can be easily adapted to trigger more than one emotion in each appraisal cycle, and it is capable of eliciting ten different types of emotions with different intensities despite having an objectively small number of fuzzy rules. Our model uses the *Pleasure* and *Arousal* dimensions internally to define emotions and mood instead of simple labels. With these dimensions, emotions and mood can be defined as vectors in which the modulus represents the intensity of the emotion, while the type of emotion is represented by the direction of this vector. Through this dimensional representation, the variables representing affective characteristics of agents (emotion and mood) can be easily adapted to obtain different emotions and moods. Therefore, our model improves the representation of emotions because it allows emotions and mood to be modified, adapted, and stored using a dimensional representation. In this work, we have also defined a fuzzy model of intensity that depends on the modulus of the emotion's vector. This model allows the agent to express emotions in the same way as humans.

One important novelty of our proposal is that our model allows the agent to express its emotional state using the same interpretation of the fuzzy emotional terms used in the language and culture of its human interlocutor. This has been achieved by defining the two-dimensional *Pleasure-Arousal* space using the results of an experiment with participants of the same cultural environment in which the agent is located. Moreover, one agent can easily adapt the expression of its emotions to other cultural environments using different *Pleasure-Arousal* spaces that can be easily incorporated to its affective component.

We are currently integrating our emotional model in the affective agent architecture *GenIA³* in order to allow the development of different behaviors depending on the agent's personality and mood. We are interested in defining a new *Pleasure-Arousal* space that is adapted to other language and/or cultural environments. This will allow agents to adapt their emotion expression to interlocutors of different cultural environments. We are also interested in analyzing the effect of adding the *Dominance* dimension to the *Pleasure-Arousal* space. We think that this additional dimension could

disambiguate the overlaps that occur between some emotions in the two-dimensional model.

A formal extension of the agent-oriented programming language agentSpeak for modelling affective agents with empathic abilities

Abstract

In this paper, we present a formalization of an extension of the syntax, the semantics, and the reasoning cycle of AgentSpeak. Our proposal modifies the AgentSpeak inference model to provide support to the development of empathic agents. The new inference model allows empathy to affect the agent's deliberation and reasoning processes to select plans according to the analyzed event and the affective state and personality of the agent. In addition, our proposal allows a software agent to simulate the distinction between self and other agents through two different event appraisal processes: the empathic appraisal process, for eliciting emotions as a response to other agents emotions, and the regular affective appraisal process for other non-empathic affective events. The empathic regulation process adapts the elicited empathic emotion based on different factors including intrapersonal (e.g., the agent's personality and affective memory) and interpersonal (e.g., the affective link between the agents) characteristics of the agent. The use of a memory of past events and the corresponding elicited emotions allows the maintaining of an affective link to support long-term interaction between agents.

5.1 Introduction

Computer systems that are oriented to the paradigm of human-machine interaction are becoming more relevant in society. Systems of this type are progressively becoming more complex and require higher level abstractions and metaphors to describe capabilities and characteristics that cannot be explained by classical lower level specifications. It is now possible to reflect the way in which human beings understand the world more faithfully through these abstractions and metaphors. Endowing software agents with affective abilities, particularly empathic abilities, is one of the metaphors that will make it possible to transform the human-machine interaction paradigm by making it more human-oriented. Empathy is a concept that has evolved over the years, whose meaning (which is still being discussed) refers to a wide range of affective and cognitive competencies that are fundamental in the development of the human being as a social being [98]. These competencies allow one "to put oneself in others' shoes" and to understand and share their feelings and thoughts, which generates a series of pro-social behaviours aimed at improving their well-being [50]. Empathy provides a medium for the establishment of affective links that support moral and social behaviour [22]. Therefore, empathy is a key element in social interactions affecting human affective states and behaviours, providing a basis for long-term relationships. In fact, social disorders, such as autism spectrum disorder, are related to a deficit in empathic abilities [95]. In any empathic interaction, we can distinguish two actors: the *target* actor, who is the person or agent that suffers the effect of an action or situation, and the *observer*, who is the person that perceives that action or situation and feels empathy for the target person.

Different theoretical approaches that try to explain the cognitive processes related to empathy have been proposed in psychology, sociology, philosophy, ethology, and neuroscience. The most recent theoretical approaches provide a conceptualization of empathic behavior from a perspective derived from appraisal theories that relate the emergence of emotions to a cognitive appraisal of an event [135, 72]. Under this perspective, empathy arises from the appraisal of the perception of a situation or an emotion in others [231, 96]. Moreover, according to several authors, empathy is affected by a regulation

process that modulates the empathic response to an event according to different factors, including intrapersonal (e.g., personality) and interpersonal (e.g., affective link with the target) factors [50, 54, 43].

For years, the inherent characteristics of human affective behavior have been the subject of research in the field of affective computing [170] with the aim of improving simulations of human behaviour and human-machine interactions [7, 148]. In recent years, there has been a growing trend towards the development of models that are capable of simulating empathic abilities [165, 236]. Most of these models use the agent-oriented paradigm. Empathic agents have been shown in multiple experiments to improve the user experience in human-machine interactions [188, 161, 159]. Other experiments, such as the one presented in [138], have also demonstrated the importance of modelling long-term interactions to improve the simulation of empathic behavior in empathic agents. In general, these models are designed to be implemented in an ad hoc manner making the agent specification programmer-dependent [31, 30]. However, an empathic agent model that is based on well-known agent-oriented languages such as AgentSpeak [179] could provide a basis for the development of agents with rational and affective abilities. This could reduce the dependency on the programmer and thus increase usability.

In this paper, we propose an empathic agent model that is designed for AgentSpeak. Our model is based on recent empathic appraisal theories to elicit empathic emotions when a software agent perceives or is aware of an emotion or a situation in other agents or humans. In addition, our empathic agent model has an *emotion regulation* process that adapts emotions to different intrapersonal and interpersonal factors. The rest of this paper is organized as follows. In Section 5.2, we perform an introduction to the theoretical frameworks supporting affective states and empathy, and we discuss the proposals made in affective computing to simulate empathic behavior. Section 5.3 introduces our AgentSpeak extension to enable the development of agents with empathic abilities. Finally, the main conclusions and some future works are presented in Section 5.4.

5.2 Related work

Human beings are emotional beings, and different affective characteristics motivate human behavior. Traditionally, the simulation of human behavior through computational systems has generally focused on practical reasoning. However, classical practical reasoning has some limitations when simulating non-rational behaviors such as affective behaviors. For years, the area of affective computing [170] has been working to design models to understand and simulate these affective behaviors [7, 148, 186, 82, 220]. These models generally include different constructs such as affective states, personality, or empathy that come from different branches of psychology, ethology, philosophy, and sociology. Affective states, which encompass emotions and moods, are complex constructs whose definition remains a subject of debate to this day. Over the years, different scientists have tried to provide an answer to the phenomenon of emotions and moods. Although emotions are generally considered to be quick responses to certain stimuli, there are different approaches that try to explain how this phenomenon arises. One of the most recognized theories is the appraisal theory [135, 72], according to which, when a stimulus is received, an appraisal process occurs, resulting in an emotion. These theories are based on the existence of different variables known as *Appraisal Variables*. Different authors differ in the number and type of *Appraisal Variables* involved in the emotion generation process. For example, K.R. Scherer [200] proposed a model in which twenty-two *Appraisal Variables* were defined, while A. Ortony, G. Clore, and A. Collins [163] proposed a more simplified model, known as OCC, with only eight *Appraisal Variables*. However, appraisal theories are not the only theories that have been developed to explain the phenomenon of emotion. Basic emotion theories relate events to specific emotions using a limited number of emotions. For example, the basic theory of emotions proposed by P. Ekman [65] uses six basic emotions: *Happiness*, *Surprise*, *Fear*, *Anger*, *Disgust*, and *Sadness*. In contrast to these basic emotion theories, the constructivist theories advocate for a more universal concept of emotions. Constructivism holds the existence of an unlimited number of emotions due to cross-cultural differences [196]. One of the most recognized constructivists is J.A. Russell who proposed the Circumplex Model of Affect

[193]. According to Russell, emotions can be expressed in a two-dimensional space that is composed of the dimensions of *Pleasure* and *Arousal*. This theory is in line with the findings of neuroscience made in recent years [174]. Other authors have proposed different dimensions to explain emotions. For example, A. Mehrabian [154] added a third dimension, called *Dominance*, that helps to disambiguate some emotions such as *Fear* and *Anger*.

On the other hand, although the differences between emotions and mood continue to divide the academic community, it is generally accepted that, in contrast to emotions, mood is not related to a specific stimulus but rather arises as a consequence of the succession of different stimuli. Mood also has a longer duration (from minutes to days) and a lower intensity than emotions [19]. The variation in mood duration is sometimes attributed to the time the mood spends in transitioning to a neutral state or equilibrium state [167]. Both the equilibrium state and the speed of transition may vary depending on the individual's affective characteristics. Moreover, it should be considered that both emotions and mood are influenced by personality [197]. Different studies over the years have shown that personality influences emotional behavior [64, 134]. Personality can make an individual more or less prone to certain emotions or moods, including complete suppression of certain emotions or moods [133]. Thus, personality acts as a factor in the regulation of affective behaviors offering an explanation for the observed differences in individual attitudes when facing the same situation [91]. One of the most recurrent theories of personality is the Five Factor Model (FFM), which is also known as the OCEAN model [152]. This model defines personality through five dimensions: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*. These five dimensions have an effect on the affective states that can be elicited in an individual. Apparently, the most evident relationships arise when relating positive affective states to *Extraversion* and when relating negative affective states to *Neuroticism* [56].

Finally, empathy is a construct that is used in different domains, such as psychology, ethology, sociology, or philosophy, to describe a variety of psychological attitudes that enable the development of social individuals. In general terms, empathy is an ability that allows humans to understand and feel the affective state of others, resulting in behavior directed toward mu-

tual understanding. Several authors consider empathy to be a fundamental psychological ability that allows the establishment and maintenance of social bonds [22, 100, 184, 157, 205]. Empathy also plays a fundamental role in our society, affecting both morality and mutual understanding. [99]. In fact, empathy appears to be the basis on which social behaviors and interpersonal interactions are founded [50, 149]. Empathy connects people by promoting relationships and collaborations through the exchange of experiences, needs, and desires. [184]. However, the definition of the term empathy continues to divide the academic community [164]. This is largely due to the fact that the concept of empathy encompasses a large number of emotional, ethical, moral, and social factors. This lack of consensus can be clearly seen in the review conducted in [49], in which the authors identified forty-three different definitions for the concept of empathy.

The modern concept of empathy is attributed to psychologist E. Titchener who translates the concept "Einfühlung" proposed by the German philosopher R. Vischer and later expanded by T. Lipps through his theory of "Einfühlung" [78, 229]. In this earlier definition, empathy was described as an innate instinct that produces a self-awareness in the experience and awareness of the target in the observer without any perspective-taking, associative, or cognitive process [230, 213]. This concept evolved over the years in different theories until the discovery of mirror neurons [187] at the end of the 20th century, which reopened the debate on the conceptualization of empathy. Mirror neurons establish a relationship between perception and action that has led to a greater understanding of how concepts such as imitation and empathy are produced in the brain and are related to emotion experiences [104, 73, 130, 62, 70]. Based on this relationship between perception and action, Preston and de Waal [176] proposed a theory of empathy known as the *Perception Action Model* (PAM). According to the PAM, when the observer perceives an emotion in the target, he/she can experience the same emotion automatically and not consciously, producing a matching of mental states between the observer and the target [175].

Nowadays, theories that use a multidimensional perspective in which empathy is embedded in a functional sense and related to high level cognitive abilities are becoming more and more relevant [226, 145]. Some of these the-

ories argue that, if empathy was an automatic process resulting from direct perception as suggested by the PAM, then humans would be constantly empathizing [96, 52]. However, there are situations in which empathy is inhibited for several reasons. This inhibition is produced by an adaptation resulting from a cognitive process known as *empathic regulation.* For instance, professionals in psychiatry are able to distance themselves from the patient's emotions [94]. F. De Vignemont and T. Singer [52] provided an explanation for this phenomenon by identifying two types of empathic regulation: empathy can be regulated by a voluntary process of *self-regulation* that improves with experience, as is the case of professionals in psychiatry; or there is an automatic and involuntary process that the authors call *implicit appraisal*, which depends on different regulation factors. Factors that influence empathy regulation have been widely discussed in academia [231, 176, 53, 51, 55]. In general, these regulation factors can be grouped into three categories: factors related to the *internal characteristics* of the observer such as personality, mood, age, life experiences, or gender; *situative context* factors such as the existence of more than one target suffering different experiences, which makes it difficult to empathize; and factors concerning the *level of relationship* between the observer and the target. This last category encompasses different interpersonal factors such as: *similarity*, which are the perceived similarities between the observer and the target (e.g., gender, personality, mood, or age); *familiarity*, which is related to the observer's previous experiences with the target in similar situations; and the *affective link*, which defines the affective relationship established between the observer and the target derived from the interactions they have over time [54].

Empathy needs to be dissociated from other social emotions. For instance, some authors have shown that empathy and Theory of Mind (or perspective-taking) correspond to different processes, recruiting different brain regions [145, 212, 125]. The ToM holds that people are capable of understanding the mental states of others thanks to a system of rules based on their own experiences [226, 94, 76]. This process allows an observer to imagine what the target is thinking or feeling. Furthermore, most theorists generally agree that, to be able to distinguish a non-empathic emotion from an empathic emotion, the observer needs to be able to maintain a sense of self as distinct

from the target [231].

New advances focus on the reasoning behind empathy inhibition as evidence that empathy is much more than responses to a stimulus [50, 231, 96, 43]. Under this perspective, empathy is described as an appraisal process that is based on other-oriented perspective-taking, self-oriented perspective-taking, and ToM to try to understand the affective situation of the target. The perspective-taking does not always require the observer to perceive the target's affective states. It is described as a process of higher level understanding of the target's situation allowing other affective experiences such as vicarious emotions [164, 10, 129, 45] (e.g., when we are happy because someone has achieved a goal). Therefore, the process of appraisal and regulation of empathy includes non-affective information such as beliefs, desires, and goals [86]. Following this premise, J.D. Wondra [231] proposes an empathic appraisal model to explain the differences observed when the emotion elicited in the observer does not correspond to the emotion observed in the target. Similarly, C. Heyes [96] presents an appraisal-based system with a process that controls the empathic response.

Empathy in software agents

The relationship of empathy with social behavior and interpersonal relationships has aroused the interest of many researchers who focuses on the areas of human behavior simulation and human-machine interaction [161, 138]. Systems that are capable of simulating empathic abilities have been used in different contexts and have proved to be more reliable and more credible, thus reducing the stress and frustration of human users and improving human-machine interaction [165]. For example, in the study conducted in [159], a robot with empathic abilities was used in an educational environment. The study determined that the empathic robot was able to elicit and maintain the social engagement of the participants of the experiment. Similarly, the experiments conducted with empathic agents in [161] showed that participants had a better perception of the agent when it displayed empathic emotions. Also, the research performed in [188] showed that agents with empathic abilities increased the involvement and sociability of participants. Therefore, consid-

ering these results, it is not surprising that in the search for improvement in the simulation of human behavior and human-machine interaction, different proposals for computer systems with empathic abilities have appeared over the years. For example, Ö.N. Yalçın and S. DiPaola [235] introduced a model of an empathic agent based on the *Russian Doll* model [53]. The Russian Doll model is a psychological theory derived from the PAM model [176] that explains high level cognitive processes. This model uses the analogy of the Russian Doll, stratifying the processes related to empathy in different layers. The inner layers represent the most primitive mechanisms of empathy related to the PAM such as mimicry or emotional contagion. The top layers represent higher level processes such as perspective-taking. Following this reasoning, Yalçın and DiPaola proposed a model that is composed of three layers. In the bottom layer, the *communication competence* contains the processes related to the recognition and the expression of emotions. The middle layer represents *emotion regulation* processes. Finally, the top layer, called *cognitive mechanisms*, includes ToM and appraisal.

Emotion recognition has traditionally played a central role in the design of empathic agents. For example the work in [181] presents a model for an empathic agent that was based on the perception of the target's emotion to simulate an empathic interaction. The system recognizes the target's emotion and makes an estimation of the empathic emotion by means of a regulation process that uses the agent's personality and mood as regulation factors. They use the mapping proposed in [153] to establish the relationship of personality and mood in a two-dimensional space that is based on *Pleasure* and *Arousal*. Another interesting approach can be seen in [31] in which an embodied virtual agent model with empathic abilities known as EMMA (Empathic MultiModal Agent) is presented. In that model, an agent is essentially composed of a reasoning module and an empathic module. The reasoning module is based on the BDI (beliefs, desires, and intentions) architecture [180]. The empathic module has a facial expression recognition system and an empathetic appraisal process that is based on the De Vignemont and Singer theory [52]. To estimate the empathic emotion, the agent attempts to internally mimic the recognized facial expression using a set of patterns called Action Units (UAs). Subsequently, the emotion is regulated in an emotion regulation

process that is based on the PAD model [154]. This process modifies the PAD components of emotions by means of a set of regulation factors that include similarity, mood, and deservedness (which represents the degree to which the target deserves or does not deserve the event). Finally, the elicited empathic emotion is translated into an Action Unit and shown in the agent's facial expression. Other models focus more on the simulation of empathy through the simulation of high-level cognitive processes. For instance, the model presented in [188], which also relies on the theory of De Vignemont and Singer to define the empathic process, focuses on the self-projection appraisal based on OCC to elicit the empathic emotion. The appraisal process is based on a set of predefined rules that define the relationship of events to desirability. The elicited emotion is adapted by using the affective link, similarity, mood, and personality in an *empathic regulation* process. Similarity is obtained by comparing the intensity and valence of both the perceived and the elicited emotions. Personality is composed of a set of thresholds for each emotion. However, these factors only affect the intensity of the emotion elicited in the empathic appraisal. Another example of the simulation of cognitive processes can be found in [161] in which a model of an embodied empathic dialogue agent is presented. The purpose of this agent was to simulate an empathic interaction with human users. The agent deduced the user's emotions through the dialogue using a perspective-taking strategy and then responded by showing the same emotion. The agent was able to regulate the intensity of the empathic emotion based on a preset degree of empathy between the agent and the user. Experiments conducted with the agent showed that participants had a better perception of the agent when it displayed empathic emotions.

Different authors have emphasized the need to consider the inclusion of long-term interactions to improve the simulation of empathic interactions [165, 139, 117]. However, not much work has been done on this topic. For instance, in [138] a model of empathy based on emotion recognition is proposed. The system is based on a robot called iCat that interacts with children while playing chess. The iCat maintains the actions that children perform in chess moves in its memory and uses this information to determine the next move taking into account the children's emotions.

From what has been explained above, it can be deduced that most of the

models of agents with empathic abilities use three fundamental processes: a *perception process*, in which an emotion or situation is perceived; an *empathic appraisal* process, in which the empathic emotion is elicited; and an *emotion regulation* process, in which the emotion is adapted to different regulation factors.

On the other hand, there is a wide range of proposals that focus on the agent-oriented programming (OAP) paradigm to develop agents with empathic abilities. Most of these proposed models are built ad hoc to simulate a certain empathic behavior, making the agent specification highly dependent on the programmer. However, the use of a commonly-used agent-oriented programming language can help to improve the development of empathic agents. In addition, these languages generally include support for processes such as perception, multi-agent communication, or rational behavior. One of the most well-known agent-oriented programming languages is AgentSpeak [179]. AgentSpeak is a language that is based on the logic programming paradigm for developing agents based on a BDI architecture. The BDI architecture provides the basis for the development of agents based on practical reasoning. Practical reasoning is an inference process through which agents evaluate and weigh their options taking into consideration the context of the practical situation in which they are involved and their knowledge about the environment. The result of this inference process is the modification, deletion, or addition of beliefs and intentions, altering the mental state of the agent. A deeper interpretation of the concept of practical reasoning allows the definition of two basic actions: deliberation and means-end reasoning. The deliberation process allows the agent to choose the state that it wants to achieve, while the means-end reasoning focuses on the actions that are required to achieve that state. The are some approaches that have used the BDI model to design agents with affective capacities [31, 6]. For example, in [85] an appraisal process based on the OCC model for eliciting emotions in BDI agents is proposed. However, that work use only one appraisal process without considering the empathic emotions.

Over the years, the AgentSpeak language has been extended by different researchers to improve its usability [224, 26, 24]. The majority of these extensions were focused on the development of Jason [25]. Jason is a platform for

the development of multi-agent systems implemented in Java whose core is an AgentSpeak interpreter. This platform enables the development of rational agents based on practical reasoning.

Despite the advantages provided by a well-known language such as AgentSpeak, there are very few proposals that use it when defining empathic agents. In [113], an empathic agent model based on AgentSpeak is presented for the resolution of conflicts of interest in interactions between agents. However, this approach focuses on utility-based functions to find solutions that are mutually acceptable, but no affective characteristic of the agents is considered in the utility functions. One of the most significant efforts to allow the use of affective agents in AgentSpeak is the *GenIA³* architecture [6]. *GenIA³* is a general-purpose architecture that extends Jason providing a platform that is easily adaptable to different theories of emotion in order to facilitate the development of affective agents. *GenIA³* adds new processes for the management of affective states, which are described in detail in Section 5.3. All of the new affective processes are grouped together to constitute the affective cycle of an agent. This affective cycle runs parallel to the rational cycle, allowing the agent to evaluate events at an affective level and consequently modify its behavior to cope with these events. However, before the work presented here, this architecture did not provide the necessary support to the development of agents with empathic abilities.

5.3 Formalization of the AgentSpeak extension to develop empathic agents

As discussed above, the use of high-level abstractions and metaphors, such as empathy, in the definition of human-oriented systems contributes to improving the experience of human-machine interactions and to make software agents more realistic when modelling human organisations. One of the most widely used paradigms proposed to model rational and affective behaviour is the agent-oriented paradigm. As stated in the previous sections, some proposals have incorporated affective characteristics into BDI-based agents; however, only few of them consider the modelization of empathic and ratio-

nal behaviour simultaneously. Moreover, in general, these proposals are based
on ad hoc specifications and do not take advantage of well-known agent lan-
guages. In this section, we present an empathic agent model based on AgentS-
peak, which is one of the best known agent-oriented programming languages.
Our model extends the syntax, semantics, and the reasoning cycle of AgentS-
peak to support the development of agents that are capable of performing
empathic interactions. We propose a new inference model that allows emo-
tions and empathy to have an effect on the practical reasoning of the agent.
Emotions and empathy will be involved in both the deliberation process when
the agent is selecting an intention to commit and in the means-end reasoning
process when the agent is deciding on the plan to follow. To do that, we have
extended the $GenIA^3$ architecture with a new syntax and semantics to facili-
tate the emergence of empathy in software agents. Our approach includes the
three fundamental processes identified in the literature on empathic agents
discussed in Section 5.2: perception, empathic appraisal, and empathic regu-
lation. The perception process is an intrinsic part of AgentSpeak. However,
we have added a new process to differentiate empathic affective events from
other affective events, the *event classification* process. This process consists
of two phases. In the first phase, an event is perceived following the process
described in [25]. In the second phase, the event is evaluated to determine
whether or not it is an affective event. This second phase defines the agent's
cognitive ability to maintain a sense of self as distinct from the target agent to
elicit a target-oriented empathic emotion. This cognitive ability is simulated
using two different appraisal processes: one for empathic events and the other
for non-empathic affective events. On the other hand, empathic appraisal is
composed of a self-projection appraisal, which evaluates the event using the
agent's own beliefs, concerns, and affective memory. Finally, the *empathic
regulation* process adapts the empathic emotion to the agent's affective char-
acteristics (e.g., mood or personality) and the knowledge that the agent has
about the target (e.g., affective link or trust level).

The rest of this section is organized as follows. Section 5.3 introduces
an example scenario. In Section 5.3, a new configuration for an empathic
agent is proposed. Section 5.3 presents the new inference model. Sections 5.3
and 5.3 introduce an extension of AgentSpeak syntax to allow the definition
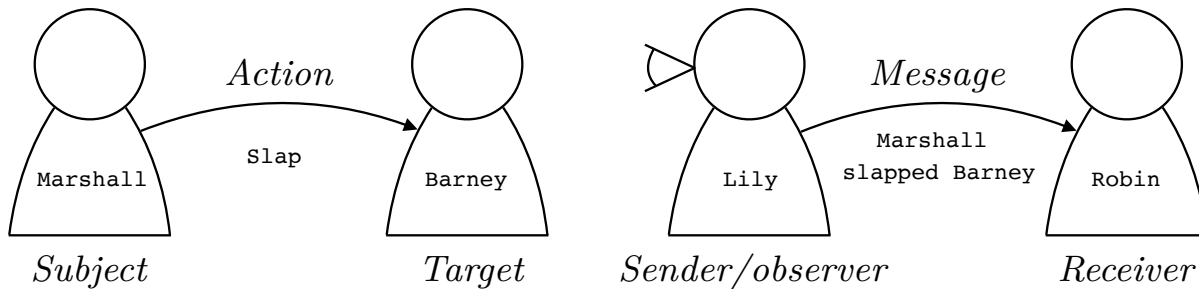
Figure 5.1: Types of agents in a social interaction

of empathic agents. Finally, Section 5.3 describes the set of semantic rules
that support the definition of the new inference model.

A simple example

From the theories of empathy previously cited, it can be assumed that, for
empathy to occur, there must be at least two actors: one who suffers in
a situation (the *target*) and another who perceives the situation and reacts
empathetically (the *observer* or empathic agent). To make the reading more
convenient, we illustrate the problem with an example. Let us assume a
more general scenario with four actors: Marshall, Lily, Barney, and Robin.
At one point in time Marshall slaps Barney in the face. This event causes
Barney to become sad. Lily, who is in the same room, sees the entire scene
and empathizes with Barney and feels sorry for what has happened to him.
Later, Lily calls Robin to tell her what happened to Barney. Robin imagines
Barny's situation and also feels sad for him.

In this environment, there are basically two interactions: first, Marshall
performs an action addressed to Barney; second, Lily perceives what has hap-
pened and performs an act of communication to inform Robin (see Figure 5.1).
Different types of agents can be identified in these interactions. On the one
hand, there are two agents that are involved in the action: Marshall, the agent
that performs the action, henceforth known as the *subject* agent; and Bar-
ney, the agent that receives the consequences of the action, henceforth known
as the *target* agent. On the other hand, in the environment there is a third
agent, Lily, who perceives the action, henceforth known as the *observer* agent.

$$\langle ag, \quad C, \quad M, \quad T, \quad s, \quad Mem, \quad Ta, \quad ast\rangle$$

$$\langle P,ps,cc,bs\rangle \; \langle I,E,A\rangle \; \langle In,Out,SI\rangle \; \langle \rho,\iota,R,Ap,\varepsilon\rangle \; \{\varepsilon,\varepsilon',...\} \; \langle Ub,Av,Cs,\sigma\rangle$$

$$\langle tr,rl,cs\rangle \; \{i,i',...\} \quad \{a,a',...\} \qquad \{p,p',...\} \; \langle te,i\rangle \; \langle Ba,Br,st\rangle$$

$$\{\langle te,i\rangle,\langle te',i'\rangle,...\} \; \{msj,msj',...\} \; \{\langle mid,i\rangle,\langle mid',i'\rangle...\}$$

$$\langle mid,id,ilf,cnt\rangle$$

(a) Agent's configuration in $GenIA^3$ [6].

$$\langle ag, \quad C, \quad M, \quad T, \quad s, \quad Mem, \quad Ta, \quad O, \quad ast\rangle$$

$$\langle P,ps,cc,bs\rangle \; \langle I,E,A\rangle \; \langle In,Out,SI\rangle \; \langle \rho,\iota,R,Ap,\varepsilon\rangle \; \{ae,ae',...\} \; \langle Ub,Av,Cs,Ae,Ee,Fe,\sigma\rangle$$

$$\langle tr,rl,cs\rangle \; \{i,i',...\} \quad \{a,a',...\} \qquad \{p,p',...\} \qquad \langle\varepsilon,av\rangle \; \langle Ba,Br,st\rangle \qquad \langle p,a\rangle$$

$$\{\langle te,i\rangle,\langle te',i'\rangle,...\} \; \{msj,msj',...\} \; \{\langle mid,i\rangle,\langle mid',i'\rangle...\} \qquad \langle te,i\rangle \; \langle p,a\rangle$$

$$\langle mid,id,ilf,cnt\rangle \qquad\qquad \{\langle p,a\rangle,\langle p',a'\rangle,...\} \quad \{\langle id,al,\sigma\rangle,\langle id',al',\sigma\rangle,...\}$$

(b) New configuration of an AgentSpeak agent including empathy. New components are shaded.

Figure 5.2: $GenIA^3$ configuration.

Finally, there is the communicative act in which two types of agents can be identified: Lily, the *observer*, who sends a message describing the action, and Robin who receives the message. According to Agent Communication Language (ACL) terminology [160], Lily will be identified as the *sender* agent and Robin will be identified as the *receiver* agent. Both agents, the *observer* and the *receiver* agents know what has happened to Barney and react to this information with an empathetic response. Therefore, in our model, we consider that both the *observer* and the *receiver* agents, are *empathic* agents.

Formalization of an empathic agent configuration

In [179], the agent-oriented language AgentSpeak semantics is defined using an operational semantics formalism. This operational semantics defines the structure and the configuration of the agent program and the transitions derived from its internal reasoning. In this section, we extend this operational

semantics to define the new AgentSpeak configuration for empathic agents.

Figure 5.2a shows the original configuration of an agent in the architecture $GenIA^3$. The shaded attributes are the ones that $GenIA^3$ added to the original configuration proposed for AgentSpeak in [24]. We have extended this configuration to add information that is relevant to the development of empathic agents. The new attributes are shaded in Figure 5.2b. In this new configuration, an agent is defined by a tuple $\langle ag, C, M, T, s, Mem, Ta, O, ast \rangle$, where:

- ag is the set of attributes that constitute the agent defined by the tuple $\langle P, ps, cc, bs \rangle$, where:

    - P is the agent's personality represented by the tuple $\langle tr, rl, cs \rangle$, where:

        - tr is a set of personality traits. Each personality trait is a value that is bounded between $[0, 1]$ representing the level that the agent has of that personality trait. For example, when using the OCEAN model, five personality traits are defined, one for each component of the OCEAN model: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*.

        - rl is a number that is bounded between $[0, 1]$ representing the agent's rationality level. The higher the rationality level, the higher the priority of plans activated in a rational BDI cycle, making the agent more rational. However, the lower the rationality level, the higher the priority of plans activated by the affective cycle, making the agent more emotional.

        - cs is the agent's set of coping strategies. These coping strategies relate affective states and beliefs to a set of intentions that will be included in the agent's agenda.

    - ps and bs represent the set of plans and the set of beliefs of the agent, respectively.

    - cc is the set of concerns of the agent that represents the motivations, standards, ideals and/or interests of the agent.

- C is the current circumstance represented by a tuple composed of: I, which is the set of intentions $\{i, i', \cdots\}$; E, which is a set of events composed of a set of tuples $\langle$triggering event te, intention $i\rangle$; and A, which is a set of actions $\{a, a', \cdots\}$.

- M are the communication parameters represented by the tuple $\langle In, Out, SI\rangle$, where SI is a set of suspended intentions composed by a set of tuples $\langle$message identifier mid, intention $i\rangle$ and In and Out represent the list of input and output messages, respectively. Each message msj is composed of the message identifier mid, the identifier of the agent which sent the message id (i.e., the *sender* agent), the illocutionary force ilf, and the message content cnt.

- T is the temporary information of the current rational cycle consisting of a tuple containing: an applicable plan ρ, a particular intention ι, the sets of relevant plans R, the set of applicable Ap plans composed of a set of plans $\{p, p', \cdots\}$, and the event ε that triggered the rational cycle represented by a tuple $\langle$triggering event te, intention $i\rangle$.

- s is the current step of the rational cycle where:

$$s \in \{\texttt{ProcMsg}, \texttt{SelEv}, \texttt{RelPl}, \texttt{AppPl}, \texttt{SelAppPl},$$
$$\texttt{AddIm}, \texttt{SelInt}, \texttt{ExcInt}, \texttt{CrlInt}\}$$

- *Mem* is the affective memory that, in the original $GenIA^3$architecture (Figure 5.2a), consists of a set of events. In our proposal (Figure 5.2b), it consists of a set of affective events ae. An affective event ae is defined as a tuple $\langle$event ε, affective value $av\rangle$, where the affective value av is an attribute that represents the emotion that the event ε produced in the agent.

- Ta represents the temporal information of the affective cycle. In the original design of $GenIA^3$ (Figure 5.2a), it is composed of: Ub, which is a tuple containing the set of beliefs that are going to be added to the belief base Ba, the set of beliefs that are going to be removed from the

belief base Br, and the identifier of the step st of the cycle in which the beliefs are going to be added or removed; Av, which is the set of appraisal variables; Cs, which is the set of coping strategies to be executed; Ae, which is the set of emotions that can be elicited by the *appraisal* process; Ee, which is the set of empathic emotions that can be triggered by the *empathic appraisal* process; Fe, the final emotion (or emotions) resulting from the *emotion selection* process; and σ, which represents the current mood of the agent. The emotions contained in Fe will be the ones considered as active in the affective cycle. Emotions and mood are represented by a tuple $\langle p, a \rangle$, where p and a are the values of *Pleasure* and *Arousal*, respectively.

– O is a new component which is added to represent the information that the empathic agent knows about other agents in the environment. In the current design, this information is composed of a set of tuples each of which corresponds to one agent in the environment. Each tuple contains the agent identifier id, the affective link al that the empathic agent has with the agent identified by id, and the id agent's mood σ. This tuple can be extended in the future to contain more knowledge about agents in the environment, such as their concerns, goals, beliefs, or trust level. The affective link is a numerical value between [-1,1] that indicates the affective link between both agents. The greater the affective link value, the greater the relationship between the agents. Negative affective links indicate enmity between the agents. The affective link can be modified due to the interactions between the agents.

– ast is the current step of the affective cycle, where:

$$ast \in \{\texttt{EvClass}, \texttt{Appr}, \texttt{EmphAppr}, \texttt{EmReg}, \texttt{EmphReg},$$
$$\texttt{EmSel}, \texttt{AffAd}, \texttt{SelCs}, \texttt{Cope}\}$$

A new rational cycle for an empathic BDI agent

The formal definition of the AgentSpeak reasoning cycle was presented in [24]. We have extended that original AgentSpeak reasoning cycle to incorpo-
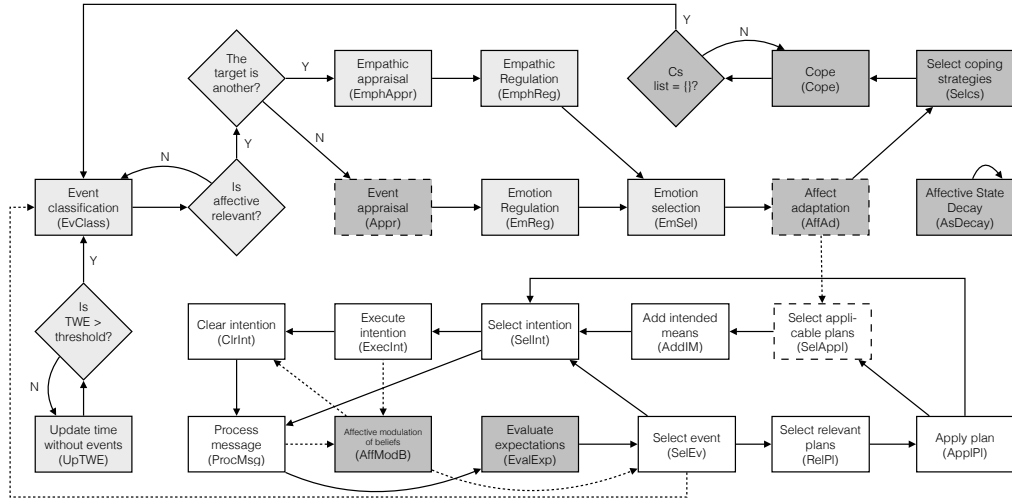
Figure 5.3: New configuration of both affective and rational cycles.

rate affective characteristics. This new reasoning cycle is represented in Figure 5.3: white boxes represent the original processes proposed for an AgentSpeak agent; dark gray shaded boxes are the original processes of the *GenIA³* architecture; dashed lines indicate processes that have been redesigned to include our extension; light gray shaded boxes represent the new processes of our extension. First, in the *process message* (`ProcMsg`) step, all of the pending messages received by the agent are processed; then, the *select event* (`SelEv`) step selects one event to process taking into account all of the perceptions, messages, and stacked events. Next, the *relevant plans* (`RelPl`) step retrieves all of the relevant plans considering the selected event; these plans are used to generate a list of applicable plans in the *apply plans* (`AppPl`) step. From this list, one applicable plan is selected in the *select applicable plan* (`SelAppl`) step. Then, in the *add intended means* (`AddIM`) step, all of the intended means are added to the set of intentions; the *select intention* (`SelInt`) step selects one intention to execute from the set. The selected intention is executed in the *execute intention* (`ExcInt`) step. Finally, the executed intention is cleared from the set of intentions in the *clear intention* (`ClrInt`) step.

This reasoning cycle was later modified by the *GenIA³* architecture to allow the development of affective agents. *GenIA³* introduced a new affective cycle that consist of five steps: the *event appraisal* (`Appr`) step in which

the selected event is appraised; the *affect adaptation* (`AffAd`) step in which the agent's mood is modified to adapt the mood according to the result of the appraisal; the *select coping strategies* step in which the coping strategies are selected according to the agent's mood; and the *cope* (`Cope`) step, in which the coping strategies are executed. In addition, *GenIA³* incorporates an *affective state decay process* (`AsDecay`) that continuously shifts the mood toward an equilibrium state to simulate real mood change in humans. The rate at which the mood moves to the equilibrium state can be set by the user. Finally, *GenIA³* modified the rational cycle to add two steps: the *affective modulation of beliefs* (`AffModB`) step in which beliefs are modified according to the agent's mood; and the *evaluate expectations* (`EvalExp`) step, which includes the possibility to add temporal expectations to agents [217].

We have redesigned the *GenIA³* architecture by adding new processes to enable the development of empathic agents. Figure 5.3 shows the new steps of both affective and rational cycles in light gray. We have introduced five new steps to the affective cycle. The first step is the *event classification* (`EvClass`) step in which an event is analyzed to determine if it is an affective event and if it can produce empathy. First of all, the process analyzes whether the event can elicit an emotion, classifying the event as affective or non-affective. Second, the process analyzes whether the affective event is directed to another agent, classifying the event as an affective empathic event or the event is directed to the agent itself, in which case, the agent will consider this event as a non-empathic affective event. This second part of the *event classification* (`EvClass`) step allows the agent to maintain a sense of self as distinct from the target. If the selected event is a non-empathic affective event, in the next step, the *emotion regulation* (`EmReg`) step adjusts the emotions produced by the appraisal according to the affective characteristics of the agent. Otherwise, if the selected event is an empathic event, the *empathic appraisal* (`EmphAppr`) process will determine a set of possible empathic emotions based on the perceptions and knowledge that the *empathic* agent has about the *target* agent. Then, the *empathic regulation* (`EmphReg`) process adapts these empathic emotions to the affective characteristics of the agent taking into account the knowledge that the *empathic* agent has about the *target* agent. Finally, in the *emotion selection* (`EmSel`) process, the final

emotion to be triggered by the empathic agent is selected. This emotion will modify the agent's mood, and will affect the selection of the applicable plan and the selection of coping strategies. In addition, we have proposed a new process named *update time without events* (`UpTWE`), which allows the development of affective behaviors such as *boredom* or *sleepiness* when the agent is not receiving events.

Extending the AgentSpeak language to identify the actors of an empathic interaction

In AgentSpeak, an agent is defined as a set of beliefs and a set of plans. Beliefs define the knowledge that the agent has about the environment. The set of beliefs of the agent is known as the *belief base*. The knowledge contained in the belief base may not necessarily be complete or accurate, since the environment may be very large and may experience changes that the agent has not perceived. On the other hand, plans contain basic actions that the agent can perform to change its environment. Plans are composed of a triggering event, a context, and a set of sequential instructions that may include: updates to the belief base, actions, or goals. Triggering events refer to the addition or deletion of beliefs or goals. In AgentSpeak, both beliefs and goals are defined as atomic formulas [179]. An atomic formula representing a belief or a goal is composed of a predicate and a set (possibly empty) of n terms of a first order logic:

$$(5.1) \qquad \qquad \texttt{predicate}(\texttt{term}_1, \texttt{term}_2, \cdots, \texttt{term}_n)$$

For example, the triggering event `time(cloudy)` is composed of the predicate '`time`' and the term '`cloudy`'.

In [224], the concept of annotations was introduced to the AgentSpeak language providing agents with the ability to identify the source of the triggering event. This concept of annotation was later reformulated by [24] to allow annotations to express properties associated with events and beliefs. The proposed syntax for defining beliefs, goals, and triggering events using annotations in AgentSpeak is:

$$(5.2) \qquad \texttt{predicate}\,(\texttt{term}_1, \texttt{term}_2, \cdots, \texttt{term}_n)\,[\texttt{a}_1, \texttt{a}_2, \cdots, \texttt{a}_m]$$

where each a_i represents the ith annotation defined as:

$$(5.3) \qquad \texttt{a}_i = \texttt{functor}_i\,\left(\texttt{term}'_{i,1}, \texttt{term}'_{i,2}, \cdots, \texttt{term}'_{i,n}\right)$$

where an atom (called functor) is followed by a number of terms (called arguments). $\texttt{term}'_{i,j}$ is the jth term of the annotation $\texttt{a}_i$.

This extension of the language provided more expressiveness to AgentSpeak allowing different types of properties to be defined. For example, we can add one annotation to an event to represent the sources of the event. The keyword that we use to represent this annotation is "source". For example, in the triggering event:

$$(5.4) \qquad \texttt{time}\,(\texttt{cloudy})\,[\texttt{source(Marshall)}]$$

the annotation $\texttt{source(Marshall)}$ indicates that the source of the triggering event is the agent Marshall.

In general, there are three possible sources for an event:

– *perceptions*, which represent the information that the agent perceives from its environment. Perceptions are represented by the annotation $\texttt{source(percept)}$.

– *mental notes*, which represent beliefs that the agent acquires or deduces by itself, such as memories or changes in the agent's state. The mental notes are expressed through the annotation $\texttt{source(self)}$.

– *communications*, which is the information that comes from another agent of the system as a consequence of a communication act. Communications are represented by the annotation $\texttt{source(agent\_id)}$ where $\texttt{agent\_id}$ is the identifier of the *sender* agent that originates the communication act.

In our proposal, we extend the language proposed in *GenIA³* with new annotations that allow the events with relevant information to be contextualized for empathy simulation. As we discussed in Section 5.2, the *empathic regulation* process is affected by a set factors that include interpersonal factors related to the *target* agent such as the affective link, concerns, goals, similarity, or trust level. Therefore, to elicit an empathic emotion, it is necessary to identify the *target* agent when an event occurs. This information can be implicit in the semantics of the event, in which case the agent may deduce the *target* of the action through an inference process, or it can be explicitly incorporated into the syntax of the triggering event. We have used this second approach because it simplifies the agent's programming by easily identifying the *subject* and the *target* of an event.

We propose an extension of the expressiveness of AgentSpeak to describe the participants in the action to which the event refers. By adding new elements it will be possible to know the agents that participate in the event when an event is received. We have extended the representation of a triggering event to explicitly include the agents involved in an action (i.e., *subject* and *target* agents) without modifying the original syntax of the atomic formula of AgentSpeak. This has been achieved using two annotations to identify the agents participating in the action represented by the triggering event: the `subject` annotation, which identifies the *subject* agent that performs the action; and the `target` annotation, which identifies the *target* agent receiving the consequences of the action. Following this definition a triggering event will be represented by the structure:

$$(5.5) \quad \begin{aligned} &\texttt{predicate}\left(\texttt{term}_1, \texttt{term}_2, \cdots, \texttt{term}_n\right) \\ &\qquad \left[\texttt{subject}(subject\_id), \texttt{target}(target\_id), \texttt{a}_3, \texttt{a}_4, \cdots, \texttt{a}_n\right] \end{aligned}$$

Where *subject_id* and *target_id* are the identifiers of the *subject* agent and the *target* agent, respectively. Note that the order in which the annotations are defined is not relevant since the annotations can be written in any order. By adding these new annotations, it is now possible to identify the agents involved in any action. Based on the example of Section 5.3, we can use

the following expression to specify the triggering event "Marshall has slapped Barney":

$$(5.6) \qquad \texttt{slap[subject(Marshall),target(Barney)]}$$

This event has a predicate "slap" and two annotations: `subject(Marshall)` and `target(Barney)`. Moreover, we propose adding an optional annotation to include the value of the interaction. The interaction value of a triggering event is an optional number that is associated with the triggering event in the range of $[-1, 1]$, which identifies if the interaction is good (positive interaction value), bad (negative interaction value), or neutral (interaction value equals 0) for the agent that receives the action. The interaction value is used to update the affective link between the agents. A positive interaction value indicates that the interaction has a positive effect on the affective state of the *target* agent, improving the affective link that the *target* agent has with the *subject* agent. A negative interaction value has a negative impact in the agent's affective state and decreases the affective link. An interaction value equal to 0 denotes that the interaction has a neutral effect on the affective state of the agent. Therefore, it has no effect on the affective link.

We are going to define the functions for calculating the values associated with the annotations introduced in our proposal. These functions will be used in Section 5.3 to formalize the new internal processes of the affective cycle of an agent. Let us define te_{annots} as the set of annotations associated to the triggering event te. The function to obtain the *subject* agent of a triggering event te is defined as:

$$(5.7) \quad \texttt{getSubject}(te) = \begin{cases} term_i, & \text{if } \exists\, a_i \in te_{annots} \mid functor_i = \text{`}\texttt{subject}\text{'} \\ & \text{and } term_i \neq ag\_id \\ \texttt{self}, & \text{if } \exists\, a_i \in te_{annots} \mid functor_i = \text{`}\texttt{subject}\text{'} \\ & \text{and } term_i = ag\_id \\ \texttt{null}, & \text{otherwise} \end{cases}$$

where a_i is the ith annotation of the set of annotations te_{annots}, *functor* is the functor of the annotation a_i, $term_i$ is the term of the annotation a_i,

'`subject`' is a keyword that identifies the *subject* annotation, and *ag_id* is
the agent's identifier. If $term_i$ is equal to the *ag_id*, it indicates that the
triggering event comes from the agent itself. If $term_i$ is not the `ag_id`, it
indicates that the triggering event comes from another agent. Finally, if there
is no '`subject`' annotation, it indicates that this triggering event comes from
an unknown subject.

Similarly, the target of a triggering event *te* is obtained through the function:

$$(5.8) \quad \texttt{getTarget}(te) = \begin{cases} term_i, & \text{if } \exists \, a_i \in te_{annots} \mid functor_i = \text{`target'} \\ \texttt{null}, & \text{otherwise} \end{cases}$$

where '`target`' is a keyword that identifies the *target* agent annotation.

Finally, to obtain the interaction value of the triggering event *te*, we define
the function:

(5.9)

$$\texttt{getInteractionValue}(te) = \begin{cases} term_i, & \text{if } \exists \, a_i \in te_{annots} \mid \\ & \quad functor_i = \text{`InteractionValue'} \\ \texttt{0}, & \text{otherwise} \end{cases}$$

where '`InteractionValue`' is a keyword that identifies the annotation
that contains the interaction value.

Extending the AgentSpeak agent configuration

The syntax of AgentSpeak was presented in [25] using the EBNF (Extended
Backus-Naur Form) notation. This syntax was later extended by the *GenIA³*
architecture to allow the development of affective agents [6, 216, 217]. In the
original EBNF syntax of *GenIA³*, an agent (`agent`) is defined by the set of initial beliefs (`init_beliefs`), concerns (`concerns`), personality (`personality`),
initial goals (`init_goals`), and plans (`plans`). We have extended this syntax to incorporate some new attributes to the agent configuration. Figure 5.4
shows the extension of the *GenIA³* agent's EBNF syntax. We have added

agent	→ init_beliefs [concerns] [personality] [others_info] init_goals plans
others_info	→ "others__:" "[" other ("," other)* "]"
other	→ ag_id ":" "[" list_attr "]"
ag_id	→ <ATOM>
list_attr	→ (attr_label ":" (<NUMBER> | <ATOM>) ("," (<NUMBER> | <ATOM>))*
attr_label	→ <ATOM>
personality	→ "personality__:" "{" list_traits ["," rat_level] ["," coping_strats] "}""."
list_traits	→ "[" trait ("," trait)* "]"
trait	→ trait_label ":" <NUMBER>
trait_label	→ <ATOM>

Figure 5.4: Simplified extension of the agent's syntax including the new extension of *GenIA³*.

mas	→ "MAS" <ID> "{" [infrastructure] [environment] [execcontrol] [w_matrix]
w_matrix	→ "w_matrix__:" w_weights ("," w_weights)*
w_weights	→ "[" trait_label ":" list_weights "]"
trait_label	→ <ATOM>
list_weights	→ "[" weight ("," weight)* "]"
weight	→ em_label ":" <NUMBER>
em_label	→ <ATOM>

Figure 5.5: Simplified extension of the MAS project syntax including the new extension of *GenIA³*.

a new attribute to represent the knowledge that the agent has about other agents (**others_knowledge**). To represent this knowledge, we use a set consisting of the agent's identity (**ag_id**) and a list of attributes associated with that agent (**list_attr**). The list of attributes is defined as a set of tuples consisting of the attribute label (**attr_label**) and its value. One of the most important attributes is the affective link that represents the affective proximity or relationship between the agents. Following the example described in Section 5.3, let us assume that agent Marshall has an affective link with agent Barney of -0.5, an affective link with agent Lily of 0.9, and has no affective link with agent Robin. To represent this knowledge, the following sentences must be added to agent Marshall's definition:

```
//Marshall agent
```

```
    //init  beliefs
    [...]
    others__:  [  Lily:  [  affective_link:  0.9  ],
              Barney:  [  affective_link:  −0.5  ]  ]
    //init  goals
    [...]
    //plans
    [...]
```

Note that, since agent Marshall does not have any affective link with agent Robin, it is not necessary to mention Robin when defining agent Marshall. Note also that the `affective_link` is not the only attribute that can be represented. The number of attributes could be increased to represent more information such as trust level or the other agents' goals, concerns, and beliefs.

On the other hand, in $GenIA^3$, the agent's personality is defined by the keyword "`personality__:`" followed by some attributes: the traits (`traits`), which are defined as a list containing a value for each personality trait; and optionally, the level of rationality (`rat_level`), which is defined as a numerical value representing how rational the agent is; and the coping strategies (`coping_strats`), which are defined as a list of plans that will be used by the agent to deal with affective events. We have modified the definition of the personality traits to make it more user friendly. Now, the personality traits are defined using a trait label (`trait_label`) followed by the numeric value for that trait. For example, let us suppose that the agent Marshall has an *Extraversion* level of 0.9 and a *Neuroticism* level of 0.1. We can express this using the new syntax as follows:

```
  //Marshall  agent
  //init  beliefs
  [...]
  personality__:  {[extraversion:  0.9,  neuroticism:  0.1]}
  others__:  [  Lily:  [  affective_link:  0.9  ]
            Barney:  [  affective_link:  −0.5  ]  ]
  //init  goals
  [...]
```

```
//plans
[...]
```

Finally, we have introduced a new attribute ω (`em_weights`) to the multi-agent system (MAS) project. ω represents a correlation matrix between the personality traits $tr \in P$ and all of the possible emotion types $t \in Ae \cup Ee$, where Ae is the set of emotions that can be elicited by the *appraisal* process and Ee is the set of empathic emotions that can be triggered by the *empathic appraisal* process. For each pair of values (tr, t), a weight p representing how the personality trait tr influences the emotion type t can be specified, but many of these weights will be zero. Therefore, the set of weights of the emotions can be viewed as the set of correlations between emotions and personality traits: the greater the correlation between the personality trait tr and the emotion type t, the greater the value of p. For example, considering that the trait of *Extraversion E* is related to positive emotions and *Neuroticism N* does not have a very high relation with positive emotions [75], for the *Happiness* emotion, the *Extraversion* weight ($\omega_{Happiness,E}$) will be greater than the weight of the *Neuroticism* for the *Happiness* emotion ($\omega_{Happiness,N}$).

To define the syntax of the ω matrix, we have proposed the use of the keyword "`w_matrix__:`" followed by one or more weights (`w_weights`). These weights are defined as a label that identifies the personality trait (`trait_label`) followed by a list of weights (`list_weights`). This list of weights is composed of one or more emotion labels (`em_label`) followed by a number that represents the correlation between that personality trait and that emotion. For instance, let us consider the example proposed in [221] in which a personality / emotion correlation matrix is defined based on the variables *Extraversion* and *Neuroticism* and the emotions *Anger* and *Sadness*. Table 5.1 shows the correlation matrix. The definition of the ω matrix based on our proposed syntax is:

```
w_matrix__:[extraversion: [Anger: 0.5, Sadness: 0.6],
            neuroticism: [Anger: 0.8, Sadness: 0.7 ]]
```

Through our extension of AgentSpeak, the programmer can clearly define both the personality of the agent and the knowledge that the agent has about other agents. On the one hand, instead of establishing one threshold for each

Table 5.1: Example of a personality / emotion correlation matrix.

	Extraversion	Neuroticism
Anger	0.5	0.8
Sadness	0.6	0.7

emotion as seen in previous approaches, our model allows the relationship
between different personality traits and emotions to be defined using the
correlation matrix ω. Therefore, just by varying the personality traits we are
defining agents with different affective behaviors as humans. On the other
hand, in our proposal, each agent can have its own affective link matrix,
thus allowing the implementation of asymmetric relationships between agents.
For instance, based on their previous interactions, Marshall has a negative
affective link with Barney, while Barney feels a positive affective link with
Marshall. This difference in the affective links between both agents represents
what happens in human societies in real life. In addition, the affective link
between two agents can dynamically vary according to the interaction value
and the emotion generated when the agents interact with each other. This
allows the simulation of affective behavior in long-term interactions between
agents.

Definitions of non-terminal syntactic symbols that have not been described
in this section are detailed in [25, 6].

Formalization of the semantic rules of the new affective cycle

We are going to formalize the transitions between the different states of the
new affective cycle presented in Section 5.3 (Figure 5.3) using the same op-
erational semantics used to define the original rational cycle of AgentSpeak
([224]). This operational semantics is described in [171] and consists of a
set of rules that define the transitions between different configurations of an
AgentSpeak agent. To make readability more convenient, we have adopted
the syntax proposed in [224]. Following this syntax, if ag is the set of at-
tributes that define an agent, we will refer to the personality component P

of ag as ag_P. We will also refer to the empathic agent whose affective cycle we are describing as *empathic* agent. In addition, we will use the label *emph_ag_id* when referring to the empathic agent identifier.

Event classification

The *event classification* (`EvClass`) step evaluates the triggering event te component of the current event T_ε to determine whether the event is affectively relevant or not using the function:

$$(5.10) \qquad \texttt{affRelEv}(te) = \begin{cases} \texttt{True}, & \text{if } \exists\, a_i \in te_{annots} \mid \\ & \quad functor_i = \text{`}\texttt{affective\_relevant}\text{'} \\ \texttt{False}, & \text{otherwise} \end{cases}$$

where te is the triggering event, te_{annots} is the set of annotations of te, a_i is the ith annotation of the set of annotations te_{annots}, and $functor_i$ represents the functor of the annotation a_i.

Furthermore, this step also performs the dissociation between events that should be evaluated at an affective level and those that should be evaluated at an empathic level allowing the distinction between the self from the *target*. As discussed in Section 5.2, this distinction between self and target is essential in the development of empathy since it allows an agent to differentiate its own emotions from emotions produced by the perception of others. For an event, we have designed four transition rules in this step. The first rule (EvClass$_1$) is applied to events that are not affectively relevant and either have nothing to do with the *empathic* agent (e.g., the target are other agents) or events whose target is the *empathic* agent but without a *subject* agent (e.g., environmental events):

(EvClass$_1$)

$$\frac{\neg\texttt{affRelEv}(te)\ \wedge\ (tg \neq emph\_ag\_id\ \vee\ (tg = emph\_ag\_id\ \wedge\ sbj \in \{\texttt{null}, \texttt{self}\}))}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{EvClass}\rangle \rightarrow \langle ag, C, M, T', s, Mem, Ta, O, \texttt{EvClass}\rangle}$$

$$\textit{where: } T_\varepsilon = \langle te, i \rangle$$
$$tg = \texttt{getTarget}(te)$$
$$sbj = \texttt{getSubject}(te)$$

If this rule is evaluated as true, the *empathic* agent will remain in the state `EvClass` waiting for a new event. For example, let us assume that Lily is the *empathic* agent. At a certain point, in time Lily perceives the event `time(cloudy)`. The function `affRelEv` evaluates the event obtaining the value false as a result because this event is not relevant for the affective cycle. The function `getTarget` also evaluates the event and returns the value `null` since there is no annotation indicating a *target* agent at the event. Therefore, the event does not have any affective impact on Lily, and the affective cycle is not started. Lily will remain in the *event classification* step waiting for a new event.

The second rule (EvClass$_2$) deals with events that are not affectively relevant but that have some level of social interaction between the *empathic* agent and the *subject* agent.

(EvClass$_2$)

$$\frac{\neg\texttt{affRelEv}(te) \ \wedge \ tg = emph\_ag\_id \ \wedge \ sbj \notin \{\texttt{null}, \texttt{self}\}}{\langle ag, C, M, T, s, \textit{Mem}, \textit{Ta}, O, \texttt{EvClass} \rangle \rightarrow \langle ag, C, M, T, s, \textit{Mem}, \textit{Ta}, O', \texttt{EvClass} \rangle}$$

$$\textit{where: } T_\varepsilon = \langle te, i \rangle$$
$$tg = \texttt{getTarget}(te)$$
$$sbj = \texttt{getSubject}(te)$$
$$iv = \texttt{getInteractionValue}(te)$$
$$O'_{al} = \texttt{updateAl}(O_{al}, iv)$$

When this rule is evaluated as true, the affective link that the *empathic* agent has with the *subject* agent is updated by the function `updateAl`. Continuing with the example, suppose that Lily (the *empathic* agent) has an affective link of 0.5 with Marshall. At a certain point in time, Marshall says hello to Lily by the event:

(Example 1)
```
hello[subject(Marshall),target(Lily),interaction_value(0.2)]
```

In this event, Marshall is the *subject* agent. As a result of this interaction, Lily will increase her affective link with Marshall by using the interaction value 0.2. As in the previous rule, Lily will remain in the *event classification* step waiting for a new event.

The last two rules refer to events that are affectively relevant and, therefore, will trigger the affective cycle. It is in these rules where the distinction between self and target occurs. This distinction is crucial because it allows an agent to differentiate its own emotions from emotions produced by the perception of others. The EvClass$_3$ rule is evaluated as true when affective events are directed towards other agents:

(EvClass$_3$)
$$\frac{\texttt{affRelEv}\big(te\big) \;\wedge\; tg \notin \big\{emph\_ag\_id, \texttt{null}\big\}}{\big\langle ag, C, M, T, s, Mem, Ta, O, \texttt{EvClass}\big\rangle \rightarrow \big\langle ag, C, M, T, s, Mem, Ta, O', \texttt{EmphAppr}\big\rangle}$$

$$
\begin{aligned}
where: \; & T_\varepsilon = \langle te, i \rangle \\
& tg = \texttt{getTarget}(te) \\
& iv = \texttt{getInteractionValue}(te) \\
& O'_{al} = \texttt{updateAl}(O_{al}, iv)
\end{aligned}
$$

If an event is affectively relevant and the target agent is another agent, the event is considered an empathic event. Therefore, the affective cycle is started, and the *empathic appraisal* step processes the empathic event. For example, when Lily receives the event:

```
        slap[subject(Marshall),target(Barney),
(Example 2)     affective_relevant,interaction_value(-0.5)]
```

*

The function `affRelEv` evaluates the event as affectively relevant and identifies that the *target* of the triggering event is Barney. Therefore, the affective cycle of agent Lily moves to the appraisal (`Appr`) step in which the event is evaluated by means of an empathic appraisal. In this appraisal step, Lily infers how Barney feels, and she generates an empathic emotion directed towards Barney.

Finally, the rule EvClass$_4$ deals with affectively relevant events that are received by the *empathic* agent:

(EvClass$_4$)
$$\frac{\texttt{affRelEv}(te) \ \wedge \ tg = emph\_ag\_id}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{EvClass}\rangle \to \langle ag, C, M, T, s, Mem, Ta, O, \texttt{Appr}\rangle}$$

$$\textit{where: } T_\varepsilon = \langle te, i\rangle$$
$$tg = \texttt{getTarget}(te)$$

When this rule is evaluated as true, the *empathic* agent performs a self-appraisal process of the event obtaining an emotion as a result. Following the previous example, when Lily receives the event:

(Example 3)
```
        slap[subject(Marshall),target(Lily),
        affective_relevant,interaction_value(-0.5)]
```

this event is directed toward Lily (i.e., Lily is the *target* agent). Therefore, Lily will evaluate this event through a self-appraisal process in the *appraisal* (`Appr`) step.

Appraisal

In the *appraisal* step, the values of the appraisal variables Av for the event ε are determined by the self-appraisal process defined by the function:

$$\texttt{Appraisal}(\varepsilon, bs, cc, Av, Mem, Ap) =$$

(5.11)
$$\{(x_{id}, x_v) \mid x_v = \texttt{DeriveAV}(x_{id}, x_v, \varepsilon, bs, cc, Mem, Ap),$$
$$\forall(x_{id}, x_v) \in Av\}$$

where bs are the agent's beliefs, cc are the agent's concerns, Av is a list of tuples (*identifier*, *value*) containing the set of appraisal variables, Mem is the affective memory, Ap are the applicable plans, x_{id} is the identifier of an appraisal variable, x_v is the value of the appraisal variable, and $\texttt{DeriveAV}$ is a function that returns the value of each appraisal variable. Note that this function calculates a different value for each appraisal variable. To determine the derived emotions from the appraisal variables, we have defined the $\texttt{DeriveEm}$ function as:

(5.12)
$$\texttt{DeriveEm}(Av) = Ae \mid Av \mapsto Ae$$

where Av represents the set of appraisal variables and Ae represents the set of appraised emotions defined as a set of tuples $\{\langle p, a \rangle, \langle p', a' \rangle, \cdots\}$ where p and a represent the *Pleasure* and *Arousal* values associated with the emotion. Therefore, this function establishes a mapping between the set of appraisal variables and emotions. There are different proposals for performing this mapping. In our approach, we have used the model proposed in [221] where a fuzzy model of appraisal based on the K.R. Scherer appraisal theory [201] is used.

The transition rule for the *appraisal* step of the affective cycle has been defined as:

(Appr$_1$)
$$\frac{True}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{Appr} \rangle \rightarrow \langle ag, C, M, T, s, Mem, Ta', O, \texttt{EmReg} \rangle}$$

$$where: \ Ta'_{Av} = \texttt{Appraisal}(T_\varepsilon, ag_{bs}, ag_{cc}, Ta_{Av}, Mem, T_{Ap})$$
$$Ta'_{Ae} = \texttt{DeriveEm}(Ta'_{Av})$$

For instance, continuing with the event of Example 3, Lily appraises the
`slap` event eliciting the emotion *sadness* with high intensity.

Once the appraisal finishes, the next step in the affective cycle is the
emotion regulation (`EmReg`) step.

Emotion regulation

The *emotion regulation* step adapts the appraised emotion based on different
factors such as the mood or personality of the agent. This regulation process
affects the probability and intensity of emotions. For example, if the affective
link between two agents is high, the probability and intensity of emotions
should be increased. The process of *emotion regulation* is established by the
`EmphRegulation` function defined as:

$$(5.13) \quad \begin{aligned} \texttt{EmRegulation}(tr, \omega, \sigma, Ae) = \{\langle p', a' \rangle \mid \langle p', a' \rangle = \varphi_1\left(tr, \omega, \sigma, \langle p, a \rangle\right), \\ \forall \langle p, a \rangle \in Ae\} \end{aligned}$$

where tr are the personality traits of the agent, ω is the personality/emo-
tion correlation matrix, σ is the agent's current mood, and Ae is the set of
appraised emotions. Since we define emotions as vectors in a Cartesian space
composed of the dimensions of *Pleasure* and *Arousal* $\langle p, a \rangle$, the function φ_1
is defined as an operation on vectors (described in more detail in [221]) that
adapts the emotion vector, represented by the tuple $\langle p, a \rangle$, using the agent's
personality and mood. Thanks to this regulation of the emotion process, each
agent can face an event in a different way depending on his/her personality
or mood.

The transition rule for this step is defined as follows:

(EmSel$_1$)

$$\frac{true}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{EmReg} \rangle \rightarrow \langle ag, C, M, T, s, Mem, Ta', O, \texttt{EmSel} \rangle}$$

$$where:\ Ta'_{Ae} = \texttt{EmRegulation}(ag_P, Ta_\sigma, Ta_{Ae})$$

Following the previous example in which Lily appraised the *sadness* emotion with high intensity, if Lily's personality has a high component of extraversion, she is more prone to positive emotions. Therefore, the *emotion regulation* process will reduce the intensity of the emotion, producing the emotion *sadness* with medium intensity.

Once the regulation of the possible emotions has been performed, the next step of the non-empathic affective cycle will be the *emotion selection* (`EmSel`) step.

Empathic appraisal

When the target of an affective event is another agent, the event is appraised by the *empathic appraisal* process. This *empathic appraisal* process is a self-projection appraisal that allows an *empathic* agent to understand another agent's emotion or situation. Similarly to the previously defined self-appraisal step, the *empathic appraisal* is defined by the formula:

$$
\begin{aligned}
\texttt{EmphAppraisal}(\varepsilon, bs, cc, Av, Mem, Ap) = \\
\{(x_{id}, x_v) \mid x_v = \texttt{DeriveAV}(x_{id}, x_v, \varepsilon, bs, cc, Mem, Ap), \\
\forall (x_{id}, x_v) \in Av\}
\end{aligned}
$$

(5.14)

where bs is the set of beliefs, cc is the set of concerns, Mem is the affective memory, Ap are the applicable plans, x_{id} is the identifier of an appraisal variable, x_v is the value of an appraisal variable, Av is a list of tuples (*identifier, value*) one for each appraisal variable, and `DeriveAV` is a function that returns the value of each appraisal variable. Note that the beliefs and concerns may be either those of the empathic agent or those of the target agent. In addition, both the number and type of appraisal variables and the function `DeriveAV` of the empathic process can be different from the self-appraisal process. Empathic emotions can also be distinct from self-appraisal and are elicited through function:

$$
\texttt{DeriveEmphEm}(Av) = Ee \mid Av \mapsto Ee
$$

(5.15)

where Av represents the set of appraisal variables and Ee represents the set of empathic emotions defined as a set of tuples $\{\langle p, a\rangle, \langle p', a'\rangle, \cdots\}$ where each p and a represent the *Pleasure* and *Arousal* values associated with the emotions. Similarly to the `DeriveEm` function, the `DeriveEmphEm` function also establishes a mapping between the set of appraisal variables and the empathic emotions.

The transition rule for the *empathic appraisal* step of the affective cycle has been defined as:

(EmphAppr_1)

$$\frac{True}{\langle ag, C, M, T, s, \textit{Mem, Ta}, O, \texttt{EmphAppr}\rangle \rightarrow \langle ag, C, M, T, s, \textit{Mem, Ta}', O, \texttt{EmphReg}\rangle}$$

$$\textit{where: } Ta'_{Av} = \texttt{EmphAppraisal}(T_\varepsilon, ag_{bs}, ag_{cc}, Ta_{Av}, Mem, T_{Ap})$$
$$Ta'_{Ee} = \texttt{DeriveEmphEm}(Ta_{Av})$$

For example, when Lily perceives the event described in Example 2, Lily evaluates the event `slap` through a self-projection appraisal and elicits the emotion *sorry for* with high intensity towards Barney.

Once the *empathic appraisal* generates the set of empathic emotions, the *empathic regulation* (`EmphReg`) process will adapt the empathic emotions using the mood and personality of the empathic agent.

Empathic regulation

The empathy regulation step adapts the empathic emotion according to different factors such as the mood Ta_σ or personality ag_P of the agent and the affective link O_{al} between the *empathic* agent and the *target* agent. This regulation process allows the empathic response to be personalized to different *target* agents. For example, if the affective link between the *empathic* agent and the *target* agent is high, the probability and intensity of the empathic emotions will be higher. The process of empathic regulation is established by the `EmphRegulation` function defined as:

(5.16)

$$\texttt{EmphRegulation}\,(tr, \omega, O, \sigma, Ee) =$$
$$\{\langle p', a' \rangle \mid \langle p', a' \rangle = \varphi_2\,(tr, \omega, O, \sigma, \langle p, a \rangle)\ \forall\, \langle p, a \rangle \in Ee\}$$

where tr are the personality traits of the agent, ω is the personality/emotion correlation matrix, O represents the knowledge that the *empathic* agent knows about the *target* agent, σ is the agent's current mood, and Ee is the set of empathic emotions. Similarly to φ_1 in Equation 5.13, φ_2 is a function that modifies the values of *Pleasure* and *Arousal* of the empathic emotion, which are represented by the tuple $\langle p, a \rangle$, using the personality, the affective link between the agents (O_{al}), and the mood (σ).

The transition rule for the empathic regulation is defined as:

(EmphReg$_1$)

$$\frac{true}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{EmphReg} \rangle \rightarrow \langle ag, C, M, T, s, Mem, Ta', O', \texttt{EmSel} \rangle}$$

$$\begin{aligned}
where:\ &Ta'_{Ee} = \texttt{EmphRegulation}(T_\varepsilon, ag_{P_{tr}}, O, Ta_\sigma, Ta_{Ee})\\
&T_\varepsilon = \langle te, i \rangle\\
&iv = \texttt{getInteractionValue}(te)\\
&O'_{al} = \texttt{updateAl}(O_{al}, iv)
\end{aligned}$$

Following the previous example, let us suppose that Lily has a low positive affective link with Barney (e.g., 0.3). Let us recall that Lily's personality makes her prone to positive emotions. In addition, when Lily perceives the event, she has a positive mood (e.g., *happiness*). The *empathic emotion regulation* process will reduce the intensity of the empathic emotion and the result will be the emotion *sorry for* with *weak* intensity.

Once the regulation of the possible empathic emotions has been performed, the following step is the *emotion selection* (`EmSel`) step.

Emotion selection

In the emotion selection step of the affective cycle, the final emotion Fe that the agent is going to simulate is selected by the function $\texttt{SelEmotion}(\varepsilon, Ae, Ee, P)$. This function has as parameters an event ε, the list of appraised emotions Ae, the list of empathic emotions Ee, and the agent personality P. The user can adapt this $\texttt{SelEmotion}$ function to different environments and situations considering the agent personality, intensity or probability of the emotions, and other affective characteristics.

The transition rule for the *emotion selection* step is defined as follows:

(EmSel$_1$)
$$\frac{True}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{EmSel} \rangle \rightarrow \langle ag, C, M, T, s, Mem', Ta', O, \texttt{AffAd} \rangle}$$

$$\begin{aligned} \textit{where:} \ \ &Ta'_{Fe} = \texttt{SelEmotion}(Ta_{Ae}, Ta_{Ee}, Ta_{\sigma}, ag_P) \\ &ae = \langle T_{\varepsilon}, Ta'_{Fe} \rangle \\ &Mem' = Mem \cup ae \end{aligned}$$

This step calculates the final emotion Ta'_{Fe}. Therefore, the affective memory Mem can be updated to store the event T_{ε} with the emotion Ta'_{Fe} resulting from this event appraisal. This mechanism allows the agent to maintain a memory of past events and the emotion that each event produced, supporting the maintaining of long-term interactions between agents.

For instance, when Lily evaluates the event of Example 3, the self-appraisal process produces two emotions as a result: *sadness* with high intensity, and *fear* with *medium* intensity. Let us suppose that after the *emotion regulation* process the intensity of those emotions is not modified. If the user has defined the $\texttt{SelEmotion}$ function to select the emotion with the highest intensity, the empathic agent will elicit the *sadness* emotion.

The next process of the empathic affective cycle is the *affect adaptation* ($\texttt{AffAd}$).

Affect adaptation

In the `AffectAdaptation` step, the mood of the agent is updated by the function `AffectAdaptation`, which is defined as follows:

(5.17)
$$\texttt{AffectAdaptation}\,(tr, \omega, \sigma, Fe) =$$
$$\langle p', a' \rangle \mid \langle p', a' \rangle = \varphi_3\,(tr, \omega, Fe, \langle p, a \rangle) \text{ and } \langle p, a \rangle \in \sigma$$

where Fe is the final emotion selected by the *emotion selection* process. φ_3 is a function that, similarly to φ_1 and φ_2 in Equations 5.13 and 5.16, modifies the mood vector, which is represented by the tuple $\langle p, a \rangle$ to adjust it to the emotion produced by the event.

(AffAd$_1$)

$$\frac{True}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{AffAd} \rangle \rightarrow \langle ag, C, M, T, s, Mem, Ta', P, O, \texttt{SelCs} \rangle}$$

where: $Ta'_\sigma = \texttt{affectAdaptation}(ag_P, Ta_\sigma, Ta_{Fe})$

Since emotions and mood are expressed as vectors in the same representation space, the selected emotion will "attract" the mood to a greater or lesser extent depending on the agent's personality. As a result, the mood of the agent will be modified. Continuing with the previous example in which the selected emotion was sadness with high intensity, let us assume that Lily's mood is *happiness* with a low intensity. The emotion will attract the mood, and, as a result, the new mood will be *sadness* with a *low* intensity.

Once the mood of the agent is updated, the next state of the affective cycle will be the *select coping strategy* process. In this process, a coping strategy from the current set of applicable coping strategies will be used to create a plan that finally allows the agent to express its empathy: in its voice tone, in its speech, in its face, . . .

Update time without events

This step checks how long it has been since the agent has not received an event. If the time exceeds a certain *threshold*, a simulated event 'Twe_event' will be triggered to elicit emotions such as boredom or sleepiness. The time without events *Twe* will be updated by the function `updateTwe()`.

(UpTwe$_1$)

$$\frac{\texttt{Twe} > \textit{threshold}}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{UpTwe} \rangle \rightarrow \langle ag, C, M, T', s, Mem, Ta', O, \texttt{EvClass} \rangle}$$

$$\begin{aligned}
\textit{where: } T'_\varepsilon &= \langle t, i \rangle \\
t &= \text{'Twe\_event'} \\
Ta'_{Av} &= \langle av_1, av_2, \cdots, av_n, Twe \rangle \\
Twe &= \texttt{updateTwe()}
\end{aligned}$$

(UpTwe$_2$)

$$\frac{\neg\texttt{Twe} > \textit{threshold}}{\langle ag, C, M, T, s, Mem, Ta, O, \texttt{UpTwe} \rangle \rightarrow \langle ag, C, M, T, s, Mem, Ta', O, \texttt{UpTwe} \rangle}$$

$$\begin{aligned}
\textit{where: } Ta'_{Av} &= \langle av_1, av_2, \cdots, av_n, Twe \rangle \\
Twe &= \texttt{updateTwe()}
\end{aligned}$$

5.4   Conclusions

Empathy is a key element in social interactions affecting human affective states and behaviours and provides a basis for long-term relationships. Several proposals have been made from the area of affective computing to simulate empathic behaviors. These proposals are generally based on theoretical models from psychology, sociology, and neuroscience. Most computational

models for simulating empathic behaviors are generally based on the agent-oriented programming paradigm. However, these proposals are designed ad hoc, generating agent programs that are highly dependent on the programmer and that are difficult to understand, share, and re-use. In this paper, we have presented a model of empathic agent that is based on the BDI language AgentSpeak. Our proposal is based on the theories of empathic appraisal and emotion regulation. We have extended the syntax, semantics, and the reasoning cycle of the AgentSpeak language to support the design of empathic agents. We have modified the affective cycle that had been incorporated to AgentSpeak by *GenIA³*. This has added new processes and functionalities to allow the elicitation of empathic emotions. We have formalized our extension using the same operational semantics formalism used by AgentSpeak and *GenIA³*. This operational semantics defines the structure and the configuration of the agent. The formalization of all of the transitions guarantees the validity of our proposal.

Our empathic model is structured around three fundamental processes: perception, empathic appraisal, and empathic regulation. After perceiving an empathic event, the empathic appraisal process elicits a set of empathic emotions. These empathic emotions are then adapted by the empathic regulation process according to the knowledge that the agent has about the *target* agent.

The new affective cycle is divided into two different appraisal processes: one for eliciting empathic emotions induced by the perception of an emotion or situation in a *target* agent; and another for eliciting emotions based on self-directed or environmental events. This division of the affective process allows the agent to be able to maintain a sense of self that is distinct from the *target* agent.

The use of the affective link representing the affective proximity between two agents allows the personalization of the empathic response based on the *target* agent. In addition, a correlation matrix represents how each personality trait of the agent influences the emotion that will be elicited. Finally, in our proposal, agents maintain a memory of past events and the emotion that the past events produced in the agent. This affective memory, combined with the possibility of establishing and modifying the affective links with other agents,

allows the simulation of personalized long-term interactions, which are crucial in affective interactions in human beings. All of these characteristics of our empathic agent model will produce software agents that are more realistic when modelling human organizations.

We are currently considering increasing the agent's knowledge of other agents by adding new structures to store the beliefs, goals, and concerns of other agents using state-of-the-art approaches of Theory of Mind. Moreover, the affective link can be complemented maintaining a level of trust in other agents. This will improve the perspective-taking process so that the agent can evaluate perceived situations using more knowledge about the *target* agent.

Towards a model of empathic pedagogical agent for educating children and teenagers on good practices in the use of social networks

JOAQUIN TAVERNER, EMILIO VIVANCOS, AND VICENTE BOTTI

Abstract

Social networks have been a revolution for our society. The rapid expansion of these networks results in more and more children and teenagers using them regularly. However, despite the fact that most social networks have privacy control systems, most young people do not use these controls because teenagers are usually unaware of the privacy risks that exist in the Internet. This inadequate use of social media can lead to different social problems such as cyber-bullying, grooming, or sexting. The best way to prevent these risks is through education. In this paper we propose an empathic pedagogical agent model for education on good practices in the use of social networks. The agent interacts empathetically with users through the social network. The user's emotion is recognized through a camera and is processed in real time to obtain the emotion. The agent analyzes the recognized emotion and the users' actions and looks for the best strategy to advice and educate the teenager in the correct use of the social network.

6.1 Introduction

Today's society is increasingly moving towards virtual environments. Video calls, instant messaging, and social networks have transformed the way we communicate and relate to others. The incorporation of these new technologies into our daily lives, has been accelerated this year by the pandemic produced by COVID-19. People find several advantages such as instant communication, the search for lost old friends, the improvement of the prospects of finding a job, or the communication in times of confinement. The democratisation of the use of the Internet is resulting in more and more children and teenagers using them. In fact, today the young people are the biggest consumers of technology and social networks [79]. However, young people are often not aware of the privacy risk they are exposed to when using the Web [3]. An inappropriate use of social networks can have negative personal, professional, and emotional consequences [144]. These consequences are aggravated when they occur in children and teenagers, as they do not have sufficient emotional capacity to deal with certain situations that may arise from the inappropriate use of social networks. Therefore, for a social transition towards virtualised environments to be effective, it is necessary to design educational models focused on teaching the correct use of the Internet. An appropriate use of the web can turn it into a pedagogical tool that allows a great dissemination of knowledge [114]. Artificial intelligence can help in this educational task through the use of intelligent agents acting as tutors in the use of social networks. Intelligent agents have already been used in education with good results [159, 207]. In particular, agents with empathic abilities, capable of recognising and showing emotions, have been successfully used in educational tasks by increasing the sociability and participation of young people [188]. Furthermore, empathic intelligent agents are perceived as much more reliable than traditional agents by children and teenagers. This advantage makes empathic agents more suitable for tutoring [165]. In this paper we propose a model of empathic pedagogical agent to be used as a tutor to train young people in the correct use of a social network.

The rest of this paper is organized as follow. In section 6.2, we discuss the problem of privacy risk in the Internet and the most important proposals

regarding pedagogy with agents and education in the use of social networks.
Section 6.3 introduces a empathetic pedagogical agent model for education
on social networks. The agent will interact with students in an affective way
by recognizing and expressing emotions. Finally, the main conclusions and
future works are presented in Section 6.4.

6.2   Related work

Social networks have revolutionised various aspects of society. They have not
only transformed the way we communicate and relate to others, but have
also affected other aspects such as politics, business, culture, or the way we
perceive reality. Social networks also allow us to meet new people and share
our ideas. However, not everything is an advantage in social networks. These
networks have become a perfect medium for the dissemination of fake news
and disinformation campaigns. Fake news try to give a biased view of many
events in order to manipulate people's opinion.

Given the high level of social exposure to which users of social networks
are subjected, there is a risk of losing privacy. In addition, an inappropriate
comment or an unfortunate photo can harm our personal image and damage
us socially and emotionally. In fact, the privacy risk has become one of the
main concerns of users of this type of network [3]. Currently, most social
networks have privacy management systems that allow users to control the
audience of their publication. However, these systems are often ignored by
users, both because they are not aware of the potential risks of exposure of a
specific publication and because users are not concerned about privacy. This
fact is further aggravated when young people are involved since they do not
have sufficient social and emotional skills to deal with certain situations that
may arise when using social networks [156].

Nowadays, the younger population is becoming the main users of social
networks. The European General Data Protection Regulation (GDPR) [110]
allows young people over 16 to transfer their personal data to be processed
by social network providers without the consent or the authorization of the
parental responsibility holder. In addition, this regulation allows EU mem-

bers to reduce this age to 13 years. Consequently, children and teenagers spend several hours a day using their smartphones. During this time, they receive hundreds of advertisements and messages that, frequently are hoaxes and fake news. This misinformation is often spread to create a negative view of everything that is different: religious belief, political ideas, sexual tendency, immigrants, people with disabilities,. . . This misinterpretation and manipulation of real events can affect the emotional development of teenagers and children, because they have less critical thinking and tools to analyse and protect themselves of fake news. Moreover, during pandemic confinement, young people have spent many hours connected to the Internet, often without parental supervision [87]. How can they filter all the messages that receive? Ballesteros and Pizaco [14] estimated that only 22% of teenagers between 14 and 16 years old have received training on how to filter the information they find on the Internet.

On the other hand, many young people use social networks as a social showcase in which they look for group acceptance: the greater the number of followers and friends, the better their perception of their status on the social network. However, most young people do not know how to handle their privacy in social network environments [210]. Therefore, when making a publication they do not consider the risk that sharing this information may suppose. This is because they do not have the necessary skills to determine what kind of information poses a risk when disseminating it. This level of public exhibition can have negative consequences. Recent studies show that the abusive and inappropriate use of social networks can lead to negative effects such as addiction, cyber-bullying, grooming, sexting, and paradoxically, social isolation [144, 178].

Pesedia: a pedagogical social network

In recent years, some proposals for assessing privacy on social networks have been made [1, 4, 29]. Most of these proposals involve the definition of metrics and mechanisms to improve trust and avoid the vulnerability of privacy in social networks [191, 218]. One of these proposals is found in [9] in which *Pesedia*, a social network designed specifically to educate in the correct use of

social networks, is presented. *Pesedia* offers an interface that is very similar
to the interfaces of current social networks like Facebook. Young people have
their own wall where they can create and share publications. They can also
send private messages or establish groups. The network also allows to establish friendship relationships and trust levels in order to teach the appropriate
privacy measures by providing the appropriate level of privacy for the publications. *Pesedia* allows the addition of plugins to increase its functionality. For
example, in [3] two mechanisms were included to assist in the decision-making
process of selecting the privacy level of a publication. The aim of these mechanisms was to alert users about the visibility of publications in order to make
them aware of the privacy risk involved when posting. These mechanisms
were based on the model of soft-paternalism, in which specific interventions
are made in the actions performed by the student without affecting his/her
freedom of choice.

Pedagogical intelligent agents

While the establishment of metrics and mechanisms has proven to be effective
in educating about good practices in social networks, they lack the interactive
experience that an intelligent agent can offer. Intelligent agents are capable of
reasoning and reacting proactively to a perceived change in the environment.
Therefore, they are able to interact with the user by reacting to his/her actions. Intelligent agents have been used in recent years to perform tutoring
tasks in educational environments improving the motivation, socialization,
understanding, and attitudes of students [188]. For example, an educational
pedagogical agent model for tutoring children on the autism spectrum is proposed in [89]. Similarly, in [16] an educational pedagogical agent model for
helping in homework is presented. That agent was developed to operate in a
social network in which students perform their homework while participating
in a common social experience.

Many proposals made in the field of affective computing focus on the study
of empathy as a mechanism to improve the human-machine interaction [165].
Empathy is a construct used in different domains such as ethology, philosophy, psychology, or neuroscience to explain different social behaviours. These

behaviours are generally focused on the ability to understand and share emotions, moods, and mental states of others, as well as the behaviours derived from this understanding, such as prosocial behaviour or altruism [49, 213]. Empathy encourages the development of emotional bonds, allowing human beings to constitute themselves as social and moral beings [119]. In fact, recent neuroscientific studies relate some social disorders, such as autism spectrum disorder or psychopathy, with a deficit in the brain areas involved in empathic processes [21]. One of the most recurrent definitions of empathy is the one proposed by M. L. Hoffman [99], in which empathy is defined as a psychological process that makes a person have feelings more similar or congruent with the other person's situation than with his/her own situation. Therefore, it is not necessary for empathic responses to be identical or approximate to the mood of the other, but according to Hoffman any emotional reaction compatible with the mood of the other can be considered empathy. According to Hoffman's definition, to simulate the processes inherent within empathy, software agents must be able to perceive and recognize emotions or moods and react accordingly by simulating a behavior appropriated to the perceived emotion or mood.

The area of artificial intelligence has evolved by developing models of agents with different affective and empathic abilities [7, 6, 216, 234]. Most of these models use a dimensional representation for emotions and mood [220]. One of the most widely used dimensional representation is the *Circumplex Model of Affect* that uses two dimensions to represent emotions: pleasure and arousal [193]. The use of this dimensional model allows the representation and simulation of affective processes such as emotion elicitation, emotional contagion, or empathy.

Empathic agents have been used in different contexts and have proved to be more reliable and more credible, thus reducing the stress and frustration of human users and improving human-machine interaction [165]. The development of agents with empathy has also been useful in the educational field. For example, the research performed in [188] used empathic agents to prevent bullying. In this experiment it was found that agents with empathic skills increased the sociability and involvement of young users. The participants played a game in which the agents recognized and showed emotions.
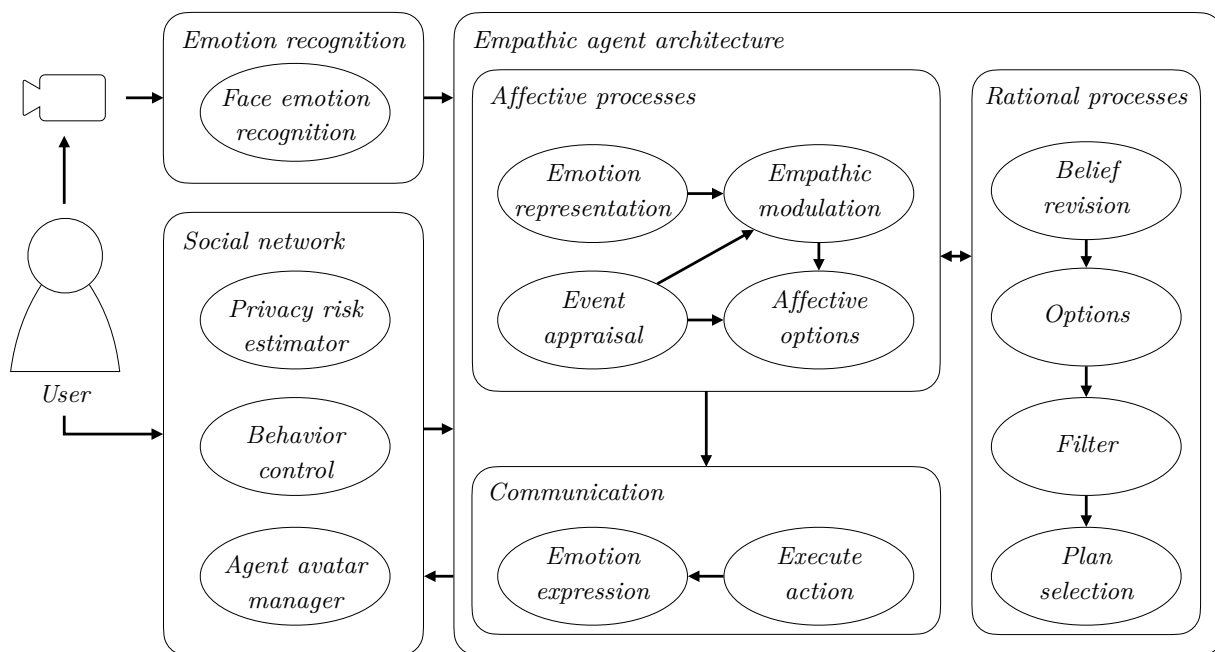
Figure 6.1: The proposed model for a empathic pedagogical agent in a social network

During the game, a series of social situations were presented in which an agent was bullied by other agents with the intention of driving the students' behaviour through empathy. Empathic agents have also been used for tutoring in educational environments. For example, in [159] a robot tutor with empathic skills was proposed. The study showed that the empathic agent improved the predisposition of young people to participate in learning experiences. In the same way, in [207] a model of a robot tutor with empathy is presented. This model is able to recognize children's emotions and interact with them through a touch screen. The results of experiments show that the empathic robot tutor was able to elicit and maintain the social engagement of the students.

Considering the promising results of previous works, where was shown that empathy increases trust and engagement, we propose to employ an empathic agent to increase the efficiency of the learning process of young people in the correct use of social networks.

6.3 Proposal

As mentioned in the previous sections, irresponsible use of social networks can compromise the user's privacy. Therefore, it is necessary to develop tools that allow young users to understand the risks to which they are exposed and to learn good practices in the use of social networks with the aim of maintaining their privacy. In this paper we present an educational pedagogical agent model with empathetic abilities. Our agent interacts with the user through the social network. The agent will monitor the user's actions on the network to analyse his/her behaviour. When the user prepares a publication, the agent analyses the privacy risk associated with the publication and recommends the best strategy for avoiding privacy risks. Following a soft-paternalism strategy, the user will be free to decide how to carry out the publication. Once the publication has been made, the agent will react according to whether the user has followed its advice or not.

In order to improve the interaction with the user, the agent must have certain social and emotional capacities to react with an appropriate behaviour. One of the best approaches to generating this social behaviour is through agents that simulate empathy. Our proposed agent uses empathy as a tool to build trust with the user improving its educational task. Through empathy, our agent is able to react appropriately according to the emotion or mood of its young interlocutor. Following the definition of empathy proposed by Hoffman, our agent will be able to recognize the emotion in the teenager and to produce a behavior according to that emotion. The model that we propose in this paper uses a real-time facial emotion recognition system. Once the emotion is recognised, the agent represents the emotion internally using a multidimensional representation model [220] and generates an empathic response in accordance with the perceived emotion. In addition, the agent will have a proactive emotional behaviour depending on the actions performed by the user in the social network.

Figure 6.1 shows the model for the interaction between the young user and the agent in the social network. The framework proposed in this work consists of three main components: the social network based on *Pesedia*, the emotion recognition system, and the empathic intelligent agent.

Pesedia

Our proposed framework uses three plugins of the *Pesedia* social network: the *privacy risk estimator*, the *behavior control*, and the *agent avatar manager*. On the one hand, the *privacy risk estimator* is responsible for estimating the potential audience for a publication. This estimation is obtained by measuring the number of potential users that can access to the publication according to the degree of privacy that the user has selected [2]. On the other hand, the *behavior control* is used to capture the interactions that the user makes within the social network, for example, when pushing a "like" button or when posting an item. Finally, the *agent avatar manager* displays the agent's facial expressions and messages on the social network window. Figure 6.2 shows a screenshot of *Pesedia* with the agent avatar interface on the right. When the young user makes a comment, the privacy risk estimator warns the user of the risk involved in such publication. In the example in Figure 6.2, the user has provided a location that could lead to a high privacy risk. The agent warns the user that the publication has a high risk and recommends to modify the audience. The final decision belongs to the user: the teenager can continue with the publication or modify the audience. Depending on his/her decision, the agent will adapt its behaviour by using different messages and emotional expressions.

The emotion recognition system

The emotion recognition system consists of a video camera and a classifier based on convolutional neural networks that is locally installed in the device used by the teenager. This system captures the image in real-time and through a cascade classifier [209] obtains the position of the user's face. Then the face image is sent to a convolutional neural network installed in the computer used by the user. This convolutional neural network obtains the most likely emotion according to the user's facial expression [74]. To avoid micro-expressions the system performs a controlled analysis using a multiple frame analysis. It is important to remark that the empathic agent can only be as good as the accuracy of the recognised emotion.

Empathic intelligent agent

Experts recommend improving the critical thinking of young people. This process involves detecting and analysing all the messages and information that the teenager receives, but frequently young people navigate the Internet without the supervision of any adult. In this kind of scenarios, an intelligent agent supervising all the teenager's activity on the Internet can provide the advice and emotional support that absent adults cannot give. Our proposal of architecture for this intelligent agent is composed of three modules: the *rational processes* module, the *affective processes* module, and the *communication* module.

Intelligent agents must be able to reason and establish plans. In our proposal the *rational processes* module is based on the BDI architecture model. A BDI agent is composed of a set of beliefs, desires, and intentions. The beliefs represent the information that the agent has about the state of the environment in which it is located, in this case the social network. The desires are objectives that the agent wants to achieve. In our model, the objectives are a set of pedagogical actions such as persuading the user to change the privacy of a message. Finally, the intentions are the objectives that the agent selects to execute them.

The *affective processes* module is composed of four processes: the *appraisal* process, the *emotion representation* process, the *empathic modulation* process, and the *selection of affective options* process. The *appraisal process* appraises any action of the young users and determines the emotion. Appraisal theories hold that when a stimulus is procured, a set of processes are triggered to evaluate that stimulus [221]. As a result of these processes an emotion is generated. In our model, when the agent perceives an action of the user in the social network, the agent evaluates this action according to its beliefs and its desires and, as result, an emotion is triggered. This emotion is represented in a two dimensional space based on the dimensions of pleasure and arousal by the *emotion representation* process. Then, the *empathic modulation* process uses the dimensional representation to approximate the emotion elicited by the appraisal process to the emotion recognized in the user's face. As a result of this process, an empathic emotion that will be more in line with the user's
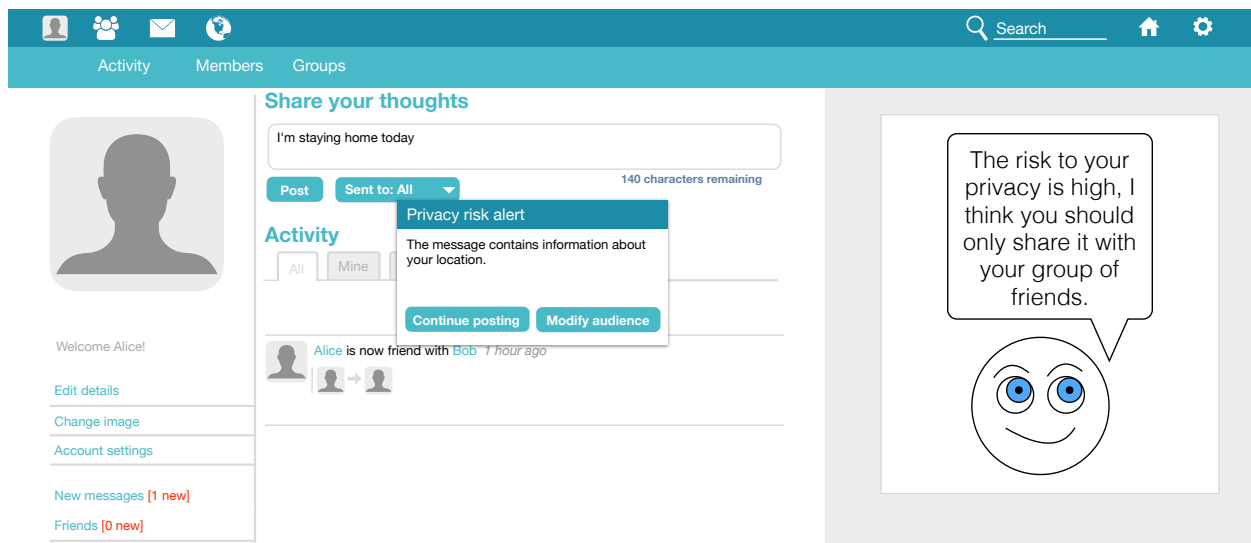
Figure 6.2: Screenshot of the *Pesedia* social network with the agent avatar interface on the right side.

emotional state is generated. Finally, this empathic emotion is used by the *selection of affective options* process to identify the applicable plans.

The last component is the *communication module.* This component is responsible for transmitting the actions that must be performed to the agent event manager. It is composed of two processes: the *execute action process* and the *emotion expression process.* The first one is in charge of executing the plans that have been selected by the plan selection process from the plans generated in the affective options and the filter processes. These plans may contain educational actions, privacy recommendations, or affective expressions. Finally, the emotion-expression process selects the facial expression that the agent must show, and sends it to the agent avatar manager (see Fig 6.2).

6.4   Conclusions and future work

Today, social networks are an essential part of many young people's lives. However, using social networks without the necessary skills represents a privacy risk. The best tool to fight the privacy risk is education. In this work we

have proposed a model of a pedagogical empathic agent for education in the use of social networks. The agent interacts with young users warning them about privacy when a publication is made. The agent interacts with the user through the social network. To make this interaction as natural as possible we have proposed the use of empathy that has been used before in teaching tasks obtaining good results.

Our empathic pedagogical agent model will be tested on *Pesedia* social network to determine if it improves the perception of the risk of privacy of young people. Further experiments will be performed to test whether the agent is able to provide young people with the necessary skills to manage privacy risk in social networks.

Part III

Discussion

Chapter 7

Discussion

Empathy is a critical component of social interactions that affects affective states and human behaviors and provides a foundation for long-term relationships. From a computational perspective, the development of systems capable of simulating abilities related to social cognition, such as empathy or Theory of Mind, requires the design of processes to identify or perceive the mental state of other agents in order to provide an appropriate empathic response. For this response of the agent to be as human-like as possible, it has to be adapted, through a process of regulation, to different affective, intrapersonal and interpersonal factors. To adapt empathic responses to these factors, it is necessary to use models for the representation of affective states that make it possible to represent, in a common space, the affective states of the agent and the affective states identified or perceived in the other agents. In this sense, theories such as the *Core Affect* theory, make it possible to represent affective states in a dimensional space and to establish relationships between the different affective states. However, as discussed in Section 2.2, proposals made to date in affective computing that used dimensional models present several disadvantages that make it difficult to extrapolate them to an empathic agent model. Therefore, this work proposes a new representation model based on a polar coordinate system composed of the pleasure and arousal dimensions, in which the affective states are represented by vectors. This model defines different types of intensity for affective states as a function of the vector module.

Furthermore, for an empathic agent to interact with human users a rep-

resentation model capable of representing and expressing emotions using the same words and expressions that humans use is needed. However, it must be considered that these words are language and culture dependent, that is, the literal translation of the terms from one language to another can produce biases or misinterpretations in the representation model. Therefore, in this thesis a new methodology based on experiments to adapt the emotion representation according to the culture and language in which it will be used is designed. This methodology is based on two experiments: the first experiment is used to study the most appropriate translations of the dimensions of pleasure and arousal in a target language; and the second experiment is used to determine the distribution of emotions along these two dimensions. The results of the first experiment with 50 Spanish participants helped to corroborate the importance of adapting the translation of emotional models by considering the different meanings that humans give to emotion labels in different cultural environments. For example, the affective category "Misery", used in the literature as the antagonistic value of Pleasure and whose closest translation into Spanish is "Misera", was not selected by the participants of the first experiment as representative of this affective category. This may be due to the fact that, in Spanish, the concept of "Miseria" is more associated with poverty than with displeasure or suffering. On the other hand, the results of the second experiment, conducted with 100 participants, helped to determine the levels of pleasure and arousal associated with the different emotions in Spanish. These experiments were performed with Portuguese and Swedish participants to adapt the model to those cultures and languages. In this case, the results allowed to highlight the differences between the three languages. For example, for some emotions, Spanish and Portuguese had a greater similarity than when compared to Swedish. These results are in line with the results obtained in previous experiments in which similarities in the conceptualization of emotions between language families were already reported. Another interesting result provided by these experiments was that there were no significant differences between the association of the dimensions of pleasure and arousal for each emotion according to gender or age.

From the results obtained in the experiments, a probabilistic model in which different probability regions are defined for each emotion was gener-

ated. In this way, the direction of the vector composed by the pleasure and arousal dimensions indicates the probability of membership of that vector to a certain affective state. This model combines probability regions with different levels of intensity, defining each affective state by an area instead of a single point. The probability areas can be easily modified to adapt the model to the cultural environment in which it is being used. In fact, through this multidimensional model of affect, an agent can have different probability models for different languages and adapt its model of representation and expression of affective states to the language and culture of its human interlocutor. To do that, two processes based on the different cultural models are established: a defuzzification process, which allows to represent words related to affective states in the space of pleasure and arousal adapted to the cultural environment; and a fuzzification process, which determines an emotion label in a specific language from a pair of pleasure and arousal values. These processes allow the agent to express its affective state and to represent the perceived affective state using the same interpretation of the fuzzy emotional terms used by its human interlocutor.

Another fundamental aspect in the development of agent models capable of simulating affective behaviors, is the way in which emotions are elicited in the agent. Based on appraisal theories, which indicate that emotions are the result of an appraisal of a perceived stimulus, a new computational appraisal model that can elicit more than one emotion in each appraisal cycle is established. This model is capable of eliciting ten different types of emotions with different intensities with a small number of fuzzy rules. To do that, this model uses five fuzzy appraisal variables to determine the candidate emotions to be elicited when perceiving an event through the *Event appraisal* process. These emotions are expressed, using vectors, in the multidimensional culturally adapted representation model. Then, through an *affect adaptation* process, the agent's mood is adapted using these emotions together with the agent's personality. As a result, different agents, with different personalities, show a different mood when they perceive the same event. To establish the influence of personality, a correlation matrix is defined. This correlation matrix can be adapted to different theories and approaches and represents how each agent's personality trait influences the emotion that will be elicited.

Once it is established how affect is represented, how affect is communicated between agents and humans, and how emotions are elicited, a formal model of an affective agent with empathic abilities is defined. This model departs from the proposals made in the field of affective computing, which generally focused on the definition of ad-hoc models. Using the generic architecture for affective BDI agents $GenIA^3$, which uses the well-known agent-oriented language AgentSpeak, the empathic agent model offers low programmer dependency. To this end, the syntax, semantics, and the reasoning cycle of AgentSpeak are extended to support the development of agents that are capable of performing empathic interactions. A new inference model is also proposed to allow emotions and empathy to have an effect on the practical reasoning of the agent. Emotions and empathy will be involved in both the deliberation process when the agent is selecting an intention to commit, and in the means-end reasoning process when the agent is deciding which plan to follow. To facilitate the emergence of empathy in software agents, the $GenIA^3$ architecture is extended with a new syntax and semantics.

In order to provide an agent with the ability to perform an empathic reasoning process based on perspective-tacking, it is necessary to clearly identify the participants involved in an empathic interaction. In the proposed model, four types of agents involved in an interaction are identified: the agent that performs the action, which is labeled as the *subject* agent; the agent that receives the consequences of the action, which is labeled as the *target* agent; and the agent that perceives the action, which is labeled as the *observer* agent. The syntax of an AgentSpeak event to incorporate the knowledge about the participants of the event is also redefined. In this way, an agent is able to reason about events that have happened to other agents and, therefore, to react empathically to these events through a process of perspective taking.

To allow the emergence of empathic interactions, the practical and affective reasoning processes of an agent are modified. In addition, three new processes for eliciting empathic emotions based on theories of empathic appraisal and emotion regulation are designed: *percepción*, *empathic appraisal*, and *empathic regulation*. After perceiving an empathic event by the *percepción* process, the *empathic appraisal* process elicits a set of empathic emotions using a self-projection approach (in which the agent evaluates the event perceived

in another agent as if the event had happened to him/her) and considering different appraisal variables. These empathic emotions are adapted by the *empathic regulation* process according to the knowledge that the agent has about the target agent. For this purpose, the *affective link* is used. The *affective link* represents the affective proximity between two agents and allows the personalization of the empathic response according to the target agent. Affective link is combined with affective memory (which stores past events together with the emotion that those events produced in the agent) for the simulation of long-term personalized interactions, which are critical in affective interactions in humans. Moreover, two types of events are distinguished: *affective events* and *empathic affective events*. Each type of event is evaluated by a different appraisal process providing the agent with the cognitive ability to distinguish the self from the others that is, according to the literature, a fundamental component in the development of cognitive empathic reasoning. All these features of this empathic agent model will lead to more realistic software agents when modeling human behavior, human organizations, and human-machine interactions.

Finally, this empathic agent model has been adapted for its use in a social network with the objective of teaching young people about the appropriate and secure use of social networks. The agent interacts with teenagers by warning them about privacy and providing guidance when posting. To make the interaction natural and believable, the agent recognizes the user's emotion and represents it using the proposed representation model adapted to its cultural environment. This emotion is used to estimate the empathic response of the agent and to adapt its behavior to the affective state of the user.

Part IV

Conclusion and future work

<div align="right">

Chapter 8

</div>

Conclusion and future work

Empathy is a fundamental component in the understanding of human social and affective behavior. However, there is still much work to be done from different areas such as psychology, sociology, or neuroscience to understand the immensity of the cognitive processes that allow the emergence of affective and social phenomena such as empathy. There are currently several theories that attempt to explain the phenomenon of empathy. The trend seems to be in line with the development of theories focused on empathic appraisal. Empathic appraisal theories suggest that empathy is generated as a result of an appraisal process together with a regulation process that allows adapting the empathic response to different social and affective factors.

The fields of computing that study the simulation of human behavior, the behavior of human social organizations, and human-machine interaction may benefit from the introduction of concepts from empathy-centered theories. Computational systems capable of simulating different aspects related to empathy contribute to improve the understanding of the factors that produce this phenomenon as well as to understand the bonds that arise in human societies. Experiments with agents capable of simulating different levels of empathy have shown that human users perceive these agents as more human-like, credible, and trustworthy. Therefore, the simulation of empathic emotional behavior allows to improve the user experience and to advance in the development of models that improve the paradigm of human-machine interaction.

However, to design this type of systems it is necessary to establish formal models that define the processes that allow the development of empathy

simulation and common representation systems of affective states for both software agents and humans. Furthermore, given the way in which humans express their emotions through labels or words, it is necessary to develop computational models capable of reasoning and simulating affective behavior by understanding and expressing affective states in the same terms as humans. However, it is necessary to consider that these terms are dependent on the cultural environment, that is, different cultures may have different ways to express the same affective state. Therefore, translating the representation models literally can lead to errors or misinterpretations affecting the simulation and the user's experience.

The work performed in this thesis represents an advance in the field of affective computing with the contribution of a model for the representation of affect and the simulation of affective and empathic behavior in multi-agent systems. The proposed model of affect representation allows, as an innovation, to represent emotions using areas in the space of pleasure and arousal instead of simple points. A emotion is represented using a vector in a polar coordinate system in which the direction of the vector indicates the type of emotion and the module of the vector determines its intensity. Through this representation it is possible to simulate different behaviors related to affective states such as the intensity of emotions or mood mitigation. In addition, having a common multidimensional representation space allows to explore affective states and adapt them to different factors (such as personality) to provide personalized responses and more realistic behaviors.

In this work, a new experiment-based methodology that allows to adapt the multidimensional model of affect representation to different cultural environments is presented. Thereby, the same agent can have different probabilistic models to adapt the emotion and express it using the same terms and in the same language of his human interlocutor, reproducing a more natural and credible simulation and avoiding possible misunderstandings or misinterpretations resulting from the different languages.

To provide agents with the ability to simulate an affective reasoning process as human-like as possible, a new computational appraisal model is defined. This model uses fuzzy logic and allows the definition of different levels of intensity for emotions. Through a fuzzification and a defuzzification processes,

the agent is able to recognize and express the affective states using the same terms and in the same language as its human interlocutor. Furthermore, using the multidimensional model of affective state representation, a formalization for mood adaptation is proposed as a function of the emotions generated by the fuzzy appraisal and the agent's personality. In this way, different agents, with different personalities, can exhibit different behaviors when perceiving the same event.

All these advances have been used to define a new agent model capable of simulating empathic abilities. Using the well-known AgentSpeak agent language, several processes have been identified and defined to simulate the cognitive processes of *affective appraisal*, *empathic appraisal*, and *empathic regulation* (among others). These processes allow agents to select the appropriate affective or empathic response when perceiving an event. The processes of *affective appraisal* and *empathic appraisal* also allow the agent to simulate the distinction between the self and the other, which is identified in the literature as a fundamental component in the development of the cognitive concept of empathy and Theory of Mind.

Therefore, this thesis represents a significant contribution to the computational representation of information related to affective states as well as to the modeling of agents with empathic and Theory of Mind abilities. However, much work remains to be done in the field of affective computing to achieve a realistic simulation of human affective behavior. Thus, some future challenges are identified. On the one hand, the representation model considers areas in a two-dimensional space formed by the pleasure and arousal components. It would be interesting to study the effect of adding more dimensions to the representation model, such as *Dominance*, as these additional dimensions may be useful to disambiguate the possible overlaps that occur between some emotions in the two-dimensional model. It would also be interesting to analyze the influence of these dimensions on social behavior by simulating it through a multi-agent system, as well as to study the influence of these dimensions across cultures. On the other hand, regarding the simulation of empathic behavior, it would be interesting to further enhance the agent's ability to use information about other agents by adding new structures to store the beliefs, goals, and concerns of other agents using state-of-the-art Theory of Mind ap-

proaches. In addition, the affective link representation can be enhanced by maintaining a level of trust in other agents. This will improve the perspective-taking process so that the agent can evaluate perceived situations using more knowledge about the *target* agent. Finally, it is necessary to develop new experiments to test whether the proposal of an empathic pedagogical agent is able to provide young people with the necessary skills to manage privacy risk in social networks.

Bibliography

[1]   Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandi-
        marte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon,
        Norman Sadeh, Florian Schaub, Manya Sleeper, et al. Nudges for pri-
        vacy and security: Understanding and assisting users' choices online.
        ACM Computing Surveys (CSUR), 50(3):1–41, 2017.

[2]   José Alemany, E del Val, J Alberola, and Ana García-Fornes. Esti-
        mation of privacy risk through centrality metrics. *Future Generation
        Computer Systems*, 82:63–76, 2018.

[3]   José Alemany, Elena del Val, J Alberola, and Ana García-Fornes. En-
        hancing the privacy risk awareness of teenagers in online social networks
        through soft-paternalism mechanisms. *International Journal of Human-
        Computer Studies*, 129:27–40, 2019.

[4]   José Alemany, Elena del Val, and Ana García-Fornes. Assisting users on
        the privacy decision-making process in an osn for educational purposes.
        In *International Conference on Practical Applications of Agents and
        Multi-Agent Systems*, pages 379–383. Springer, 2020.

[5]   Bexy Alfonso, Emilio Vivancos, and Vicente Botti. An open arquitec-
        ture for affective traits in a BDI agent. *Proc. of the 6th ECTA 2014.
        Part of the 6th IJCCI 2014*, pages 320–325, 2014.

[6]   Bexy Alfonso, Emilio Vivancos, and Vicente Botti. Toward formal mod-
        eling of affective agents in a BDI architecture. *ACM Trans. on Internet
        Technology*, 17(1):5, 2017.

[7]   Bexy Alfonso, Emilio Vivancos, Vicente Botti, and Penélope Hernández. Building emotional agents for strategic decision making. In *Proc. of the Int. Conf. on Agents and Artificial Intelligence*, volume 2, pages 390–397. SCITEPRESS, 2015.

[8]   Selim Aren and Hatice Nayman Hamamci. Relationship between risk aversion, risky investment intention, investment choices: Impact of personality traits and emotion. *Kybernetes*, 2020.

[9]   Estefanía Argente, Emilio Vivancos, José Alemany, and Ana García-Fornes. Educating in privacy in the use of social networks. *Education in the Knowledge Society*, 18(2):107–126, 2017.

[10]   Paul J Argott, Dawn Buffington Townsend, and Claire L Poulson. Acquisition and generalization of complex empathetic responses among children with autism. *Behavior analysis in practice*, 10(2):107–117, 2017.

[11]   Aladdin Ayesh and William Blewitt. Models for computational emotions from psychological theories using type-ii fuzzy logic. *Cognitive Computation*, 7(3):309–332, 2015.

[12]   Bab Dictionary. Bab English–Swedish Dictionary. https://sv.bab.la, 2021 (accessed September 17, 2021).

[13]   Kaveh Bakhtiyari and Hafizah Husain. Fuzzy model of dominance emotions in affective computing. *Neural Computing and Applications*, 25(6):1467–1477, 2014.

[14]   Juan Carlos Ballesteros and Laura Picazo. *ICTs and their influence on adolescent socialization. (in Spanish)*. Madrid: Centro Reina Sofia sobre Adolescencia y Juventud Fundacion de Ayuda a la Drogadicción, 2018.

[15]   Lisa Feldman Barrett, James Gross, Tamlin Conner Christensen, and Michael Benvenuto. Knowing what you're feeling and knowing what to

do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, 15(6):713–724, 2001.

[16] María Lucía Barrón-Estrada, Ramón Zatarain-Cabada, Yasmín Hernández Pérez, et al. An intelligent and affective tutoring system within a social network for learning mathematics. In *Ibero-American Conference on Artificial Intelligence*, pages 651–661. Springer, 2012.

[17] Erdenebileg Batbaatar, Meijing Li, and Keun Ho Ryu. Semantic-emotion neural network for emotion recognition from text. *IEEE Access*, 7:111866–111878, 2019.

[18] Edward Batschelet. Circular statistics in biology. *ACADEMIC PRESS, 111 FIFTH AVE., NEW YORK, NY 10003, 1981, 388*, 1981.

[19] Christopher Beedie, Peter Terry, and Andrew Lane. Distinctions between emotion and mood. *Cognition & Emotion*, 19(6):847–878, 2005.

[20] Reda Belaiche, Yu Liu, Cyrille Migniot, Dominique Ginhac, and Fan Yang. Cost-effective cnns for real-time micro-expression recognition. *Applied Sciences*, 10(14):4959, 2020.

[21] Robert James R Blair. Fine cuts of empathy and the amygdala: dissociable deficits in psychopathy and autism. *Quarterly journal of experimental psychology*, 61(1):157–170, 2008.

[22] Paul Bloom. Empathy and its discontents. *Trends in cognitive sciences*, 21(1):24–31, 2017.

[23] Michael Boiger, Eva Ceulemans, Jozefien De Leersnyder, Yukiko Uchida, Vinai Norasakkunkit, and Batja Mesquita. Beyond essentialism: Cultural differences in emotions revisited. *Emotion*, 18(8):1142, 2018.

[24] Rafael H Bordini and Jomi F Hübner. A java-based interpreter for an extended version of agentspeak. *University of Durham, Universidade Regional de Blumenau*, 256, 2007.

[25] Rafael H Bordini, Jomi Fred Hübner, and Michael Wooldridge. *Programming multi-agent systems in AgentSpeak using Jason*, volume 8. John Wiley & Sons, 2007.

[26] Rafael H Bordini and Alvaro F Moreira. Proving BDI properties of agent-oriented programming languages: The asymmetry thesis principles in agentspeak. *Annals of Mathematics and Artificial Intelligence*, 42(1):197–226, 2004.

[27] Jeroen Borghuis, Wiebke Bleidorn, Klaas Sijtsma, Susan Branje, Wim HJ Meeus, and Jaap JA Denissen. Longitudinal associations between trait neuroticism and negative daily experiences in adolescence. *Journal of Personality and Social Psychology*, Ahead of print (2019), 2019.

[28] Vicent Botti, Andrea Omicini, Stefano Mariani, and Vicente Julian. *Multi-agent systems*. MDPI-Multidisciplinary Digital Publishing Institute, 2019.

[29] Víctor Botti-Cebriá, Elena del Val, and Ana García-Fornes. Automatic detection of sensitive information in educative social networks. In *Conference on Complex, Intelligent, and Software Intensive Systems*, pages 184–194. Springer, 2020.

[30] Hana Boukricha and Ipke Wachsmuth. Modeling empathy for a virtual human: how, when and to what extent? In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1135–1136, 2011.

[31] Hana Boukricha, Ipke Wachsmuth, Maria Nella Carminati, and Pia Knoeferle. A computational model of empathy: Empirical evaluation. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 1–6. IEEE, 2013.

[32] Hana Boukricha, Ipke Wachsmuth, Maria Nella Carminati, and Pia Knoeferle. Empathy and its modulation in a virtual human. In *Annual Conf. on Artificial Intelligence*, pages 25–36. Springer, 2013.

[33]   Muriel Bouyer, Sophie Bagdassarian, Sveti Chaabanne, and Etienne Mullet. Personality correlates of risk perception. *Risk analysis*, 21(3):457–466, 2001.

[34]   Paul H. Bucci, X Laura Cang, Hailey Mah, Laura Rodgers, and Karon E. MacLean. Real emotions don't stand still: Toward ecologically viable representation of affective interaction. In *8th Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, pages 1–7. IEEE, 2019.

[35]   Sathya Bursic, Giuseppe Boccignone, Alfio Ferrara, Alessandro D'Amelio, and Raffaella Lanzarotti. Improving the accuracy of automatic facial expression recognition in speaking subjects with deep learning. *Applied Sciences*, 10(11):4002, 2020.

[36]   Davide Buttarazzi, Giuseppe Pandolfo, and Giovanni C Porzio. A boxplot for circular data. *Biometrics*, 74(4):1492–1501, 2018.

[37]   Linqin Cai, Binbin Liu, Jimin Yu, and Jianrong Zhang. Human behaviors modeling in multi-agent virtual environment. *Multimedia Tools and Applications*, 76(4):5851–5871, 2017.

[38]   Sergio Castellanos, Luis-Felipe Rodríguez, Luis A Castro, and Joaquín Pérez. Cognitive modulation of appraisal variables in the emotion process of autonomous agents. In *Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pages 102–108. IEEE, 2017.

[39]   Sergio Castellanos, Luis-Felipe Rodríguez, and J Octavio Gutierrez-Garcia. A mechanism for biasing the appraisal process in affective agents. *Cognitive Systems Research*, 58:351–365, 2019.

[40]   Ana Paula Chaves and Marco Aurelio Gerosa. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37(8):729–758, 2021.

[41]   Marcello Cirillo, Lars Karlsson, and Alessandro Saffiotti. Human-aware task planning: an application to mobile robots. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 1(2):1–26, 2010.

[42]   Caitlyn Clabaugh and Maja Matarić. Robots for the people, by the people: Personalizing human-machine interaction. *Science robotics*, 3(21):eaat7451, 2018.

[43]   Malissa A Clark, Melissa M Robertson, and Stephen Young. "I feel your pain": A critical review of organizational research on empathy. *Journal of Organizational Behavior*, 40(2):166–192, 2019.

[44]   Giovanna Colombetti. The embodied and situated nature of moods. *Philosophia*, 45(4):1437–1451, 2017.

[45]   Amy Coplan and Peter Goldie. *Empathy: Philosophical and psychological perspectives*. Oxford University Press, 2011.

[46]   Daniel T Cordaro, Rui Sun, Dacher Keltner, Shanmukh Kamble, Niranjan Huddar, and Galen McNeil. Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, 18(1):75, 2018.

[47]   Natalia Criado, Estefania Argente, Pablo Noriega, and Vicente Botti. Human-inspired model for norm compliance decision making. *Information Sciences*, 245:218–239, 2013.

[48]   Ana Carla Crispim, Roberto Moraes Cruz, Cassandra Melo Oliveira, and Aline Battisti Archer. O afeto sob a perspectiva do circumplexo: evidências de validade de construto. *Avaliação Psicológica*, 16(2):145–152, 2017.

[49]   Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. Empathy: A review of the concept. *Emotion Review*, 8(2):144–153, 2016.

[50]   Mark H Davis. *Empathy: A social psychological approach*. Routledge, 2018.

[51]   Mark H Davis, Carol Luce, and Stephen J Kraus. The heritability of characteristics associated with dispositional empathy. *Journal of personality*, 62(3):369–391, 1994.

[52]   Frederique De Vignemont and Tania Singer. The empathic brain: how, when and why? *Trends in cognitive sciences*, 10(10):435–441, 2006.

[53]   Frans BM De Waal. The 'russian doll'model of empathy and imitation. *On being moved: From mirror neurons to empathy*, pages 35–48, 2007.

[54]   Jean Decety and Philip L Jackson. The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3(2):71–100, 2004.

[55]   Jean Decety and Claus Lamm. Human empathy through the lens of social neuroscience. *The scientific World journal*, 6:1146–1163, 2006.

[56]   Douglas Derryberry and Marjorie A Reed. Temperament and attention: Orienting toward and away from positive and negative signals. *Journal of Personality and Social Psychology*, 66(6):1128–1139, 1994.

[57]   Sandrine Detandt, Christophe Leys, and Ariane Bazan. A french translation of the pleasure arousal dominance (PAD) semantic differential scale for the measure of affect and drive. *Psychologica Belgica*, 57(1):17, 2017.

[58]   Cambridge Dictionary. Cambridge English–Spanish Dictionary. https://dictionary.cambridge.org, 2019 (accessed March 10, 2019).

[59]   Collins Dictionary. Collins English–Spanish Dictionary. https://www.collinsdictionary.com, 2019 (accessed March 10, 2019).

[60]   Oxford Dictionary. Oxford English–Spanish Dictionary. https://es.oxforddictionaries.com, 2019 (accessed March 10, 2019).

[61]   Wordreference Dictionary. Wordreference English–Spanish Dictionary. https://www.wordreference.com, 2019 (accessed March 10, 2019).

[62] Ulf Dimberg and Monika Thunberg. Empathy, emotional contagion, and rapid facial reactions to angry and happy facial expressions. *PsyCh Journal*, 1(2):118–127, 2012.

[63] Jiaxuan Du, Jiali Huang, Yuanyuan An, and Wei Xu. The relationship between stress and negative emotion: The mediating role of rumination. *Clin Res*, 4(1):1–5, 2018.

[64] Panteleimon Ekkekakis. Affect, mood, and emotion. In *Measurement in Sport and Exercise Psychology*, pages 321–1332. Human Kinetics, 2012.

[65] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[66] Paul Ed Ekman and Richard J Davidson. *The nature of emotion: Fundamental questions.* Oxford University Press, 1994.

[67] Magy Seif El-Nasr, John Yen, and Thomas R Ioerger. Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257, 2000.

[68] Hillary Anger Elfenbein. 7 emotion in organizations: a review and theoretical integration. *Academy of management annals*, 1(1):315–386, 2007.

[69] Nacer Farajzadeh and Mahdi Hashemzadeh. Exemplar-based facial expression recognition. *Information Sciences*, 460-461:318 – 330, 2018.

[70] PF Ferrari, M Gerbella, G Coudé, and S Rozzi. Two different mirror neuron networks: the sensorimotor (hand) and limbic (face) pathways. *Neuroscience*, 358:300–315, 2017.

[71] Joseph P Forgas. Mood effects on cognition: Affective influences on the content and process of information processing and behavior. *Emotions and affect in human factors and human-computer interaction*, pages 89–122, 2017.

[72] Nico H Frijda, Peter Kuipers, and Elisabeth Ter Schure. Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, 57(2):212, 1989.

[73] Chris D Frith and Uta Frith. The neural basis of mentalizing. *Neuron*, 50(4):531–534, 2006.

[74] José Manuel Fuentes, Joaquin Taverner, Jaime Andres Rincon, and Vicente Botti. Towards a classifier to recognize emotions using voice to improve recommendations. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 218–225. Springer, 2020.

[75] Desire Furnes, Hege Berg, Rachel LC Mitchell, and Silke Paulmann. Exploring the effects of personality traits on the perception of emotions from prosody. *Frontiers in Psychology*, 10:184, 2019.

[76] Helen L Gallagher and Christopher D Frith. Functional imaging of 'theory of mind'. *Trends in cognitive sciences*, 7(2):77–83, 2003.

[77] Shaun Gallagher. The new hybrids: Continuing debates on social perception. *Consciousness and cognition*, 36:452–465, 2015.

[78] Joanna Ganczarek, Thomas Hünefeldt, and Marta Olivetti Belardinelli. From "einfühlung" to empathy: exploring the relationship between aesthetic and interpersonal experience. *Cognitive processing*, 19(2):141–145, 2018.

[79] Francisco J. García-Peñalvo and Nicholas Alfred Kearney. Networked youth research for empowerment in digital society: The wyred project. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, page 3–9. Association for Computing Machinery, 2016.

[80] Bertram Gawronski, Paul Conway, Joel Armstrong, Rebecca Friesdorf, and Mandy Hütter. Effects of incidental emotions on moral dilemma judgments: An analysis using the cni model. *Emotion*, 18(7):989, 2018.

[81] Patrick Gebhard. ALMA: a layered model of affect. In *Proc.gs of the 4th Int. Joint Conf. on Autonomous Agents and Multiagent Systems*, pages 29–36. ACM, 2005.

[82] Patrick Gebhard and Kerstin H Kipp. Are computer-generated emotions and moods plausible to humans? In *Int. Workshop on Intelligent Virtual Agents*, pages 343–356. Springer, 2006.

[83] Michael Gelfond and Vladimir Lifschitz. Action languages. *Computer and Information Science*, 3(16), 1998.

[84] Gonzalo Blázquez Gil, Antonio Berlanga de Jesús, and José M Molina Lopéz. Combining machine learning techniques and natural language processing to infer emotions using spanish twitter corpus. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 149–157. Springer, 2013.

[85] Joao Gluz and Patricia A Jaques. A probabilistic formalization of the appraisal for the occ event-based emotions. *Journal of Artificial Intelligence Research*, 58:627–664, 2017.

[86] Alvin Goldman. Two routes to empathy. In *Empathy: Philosophical and psychological perspectives*, pages 31–44. Oxford University Press Oxford, 2011.

[87] José Gómez-Galán, José Ángel Martínez-López, Cristina Lázaro-Pérez, and José Luis Sarasola Sánchez-Serrano. Social networks consumption and addiction in college students during the covid-19 pandemic: Educational approach to responsible use. *Sustainability*, 12(18):7737, 2020.

[88] Jonathan Gratch and Stacy Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.

[89] Beate Grawemeyer, Hilary Johnson, Mark Brosnan, Emma Ashwin, and Laura Benton. Developing an embodied pedagogical agent with and for young people with autism spectrum disorder. In *International*

Conference on Intelligent Tutoring Systems, pages 262–267. Springer, 2012.

[90] Elizabeth K Gray, David Watson, R Payne, and C Cooper. Emotion, mood, and temperament: Similarities, differences, and a synthesis. In *Emotions at work: Theory, research and applications for management*, pages 21–43. John Wiley & Sons Chichester, UK, 2001.

[91] James J Gross. *Emotion regulation: Conceptual and empirical foundations.* The Guilford Press, 2014.

[92] Ayesha Hakim, Stephen Marsland, and Hans W Guesgen. Statistical modelling of complex emotions using mixture of von mises distributions. In *2013 Humaine Association Conf. on Affective Computing and Intelligent Interaction*, pages 517–522. IEEE, 2013.

[93] Eddie Harmon-Jones, Cindy Harmon-Jones, and Elizabeth Summerell. On the importance of both dimensional and discrete models of emotion. *Behavioral sciences*, 7(4):66, 2017.

[94] Grit Hein and Tania Singer. I feel how you feel but not always: the empathic brain and its modulation. *Current opinion in neurobiology*, 18(2):153–158, 2008.

[95] Cecilia Heyes. Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, 34(4):575–583, 2010.

[96] Cecilia Heyes. Empathy is not in our genes. *Neuroscience & Biobehavioral Reviews*, 95:499–507, 2018.

[97] Katie Hoemann, Fei Xu, and Lisa Feldman Barrett. Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental psychology*, 55(9):1830, 2019.

[98] Martin L Hoffman. Interaction of affect and cognition in empathy. In *Emotions, cognition, and behavior*, pages 103–131. Cambridge University Press, 1984.

[99] Martin L Hoffman. *Empathy and moral development: Implications for caring and justice.* Cambridge University Press, 2001.

[100] Martin L Hoffman. Empathy and prosocial behavior. *Handbook of emotions*, 3:440–455, 2008.

[101] Ying-yi Hong, Michael W Morris, Chi-yue Chiu, and Veronica Benet-Martinez. Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7):709, 2000.

[102] Joshua T Horwood and Aubrey B Poore. Gauss von mises distribution for improved uncertainty realism in space situational awareness. *SIAM/ASA Journal on uncertainty Quantification*, 2(1):276–304, 2014.

[103] M Shamim Hossain and Ghulam Muhammad. Emotion recognition using secure edge and cloud computing. *Information Sciences*, 504:589–601, 2019.

[104] Marco Iacoboni. Imitation, empathy, and mirror neurons. *Annual review of psychology*, 60:653–670, 2009.

[105] Infopedia Dictionary. Infopedia English–Portuguese Dictionary. https://www.infopedia.pt, 2021 (accessed September 17, 2021).

[106] Leonid Ivonin, Huang-Ming Chang, Wei Chen, Matthias Rauterberg, et al. A new representation of emotion in affective computing. In *Proc. of Int. Conf. on affective computing and intelligent interaction*, pages 337–343, 2012.

[107] Joshua Conrad Jackson, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray, and Kristen A Lindquist. Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472):1517–1522, 2019.

[108] Shikha Jain and Krishna Asawa. EMIA: emotion model for intelligent agent. *Journal of Intelligent Systems*, 24(4):449–465, 2015.

[109] S Rao Jammalamadaka and Ambar Sengupta. *Topics in circular statistics*, volume 5. World Scientific, 2001.

[110] Lina Jasmontaite and Paul De Hert. The eu, children under 13 years, and parental consent: a human rights analysis of a new, age-based bright-line for the protection of children on the internet. *International Data Privacy Law*, 5(1):20–33, 2015.

[111] Felix Jimenez, Tomohiro Yoshikawa, Takeshi Furuhashi, and Masayoshi Kanoh. An emotional expression model for educational-support robots. *Journal of Artificial Intelligence and Soft Computing Research*, 5, 2015.

[112] George H John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.

[113] Timotheus Kampik, Juan Carlos Nieves, and Helena Lindgren. Empathic autonomous agents. In *International Workshop on Engineering Multi-Agent Systems*, pages 181–201. Springer, 2018.

[114] Hasan Karal, Mehmet Kokoc, and Ozlem Cakir. Impact of the educational use of facebook group on the high school students' proper usage of language. *Education and Information Technologies*, 22(2):677–695, 2017.

[115] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, and Mo Nours Arab. Human-computer interaction: Overview on state of the art. *International journal on smart sensing and intelligent systems*, 1(1), 2017.

[116] Charalampos Karyotis, Faiyaz Doctor, Rahat Iqbal, Anne James, and Victor Chang. A fuzzy computational model of emotion for cloud based sentiment analysis. *Information Sciences*, 433:448–463, 2018.

[117] Zerrin Kasap, Maher Ben Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann. Making them remember—emotional virtual characters with memory. *IEEE Computer Graphics and Applications*, 29(2):20–29, 2009.

[118] Christina Katsimerou, Judith A Redi, and Ingrid Heynderickx. A computational model for mood recognition. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 122–133. Springer, 2014.

[119] Antti Kauppinen. Empathy, emotion regulation, and moral judgment. *Oxford University Press*, 2014.

[120] Anna Kåver. *KBT i utveckling: en introduktion till kognitiv beteendeterapi.* Natur och kultur, 2006.

[121] Abe Kazemzadeh. Using interval type-2 fuzzy logic to translate emotion words from Spanish to English. In *Int. Conf. on Fuzzy Systems*, pages 1–8. IEEE, 2010.

[122] Abe Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan. Fuzzy logic models for the meaning of emotion words. *IEEE Computational Intelligence*, 8(2):34–49, 2013.

[123] Abe Kazemzadeh, Sungbok Lee, and Shrikanth S Narayanan. An interval type-2 fuzzy logic system to translate between emotion-related vocabularies. In *Ninth Annual Conference of the International Speech Communication Association*, 2008.

[124] Theodore D Kemper. *Elementary Forms of Social Relations: Status, power and reference groups.* Routledge, 2016.

[125] Olga M Klimecki, Matthieu Vétois, and David Sander. The impact of empathy and perspective-taking instructions on proponents and opponents of immigration. *Humanities and Social Sciences Communications*, 7(1):1–12, 2020.

[126] George Klir and Bo Yuan. *Fuzzy sets and fuzzy logic*, volume 4. Prentice Hall New Jersey, 1995.

[127] Igor Knez and Staffan Hygge. The circumplex structure of affect: A swedish version. *Scandinavian Journal of Psychology*, 42(5):389–398, 2001.

[128] Monica La Mura and Patrizia Lamberti. Human-machine interaction personalization: a review on gender and emotion recognition through speech analysis. In *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*, pages 319–323. IEEE, 2020.

[129] Claus Lamm, C Daniel Batson, and Jean Decety. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of cognitive neuroscience*, 19(1):42–58, 2007.

[130] Claus Lamm, Markus Rütgen, and Isabella C Wagner. Imaging empathy and prosocial emotions. *Neuroscience letters*, 693:49–53, 2019.

[131] Agnieszka Landowska. Towards new mappings between emotion representation models. *Applied Sciences*, 8(2):274, 2018.

[132] Randy J Larsen and Edward Diener. Promises and problems with the circumplex model of emotion. *In M. S. Clark (Ed.), Emotion (pp. 25–59)*, 1992.

[133] Randy J Larsen and Timothy Ketelaar. Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology*, 61(1):132, 1991.

[134] Jennifer C Lay, Denis Gerstorf, Stacey B Scott, Theresa Pauly, and Christiane A Hoppmann. Neuroticism and extraversion magnify discrepancies between retrospective and concurrent affect reports. *Journal of Personality*, 85(6):817–829, 2017.

[135] Richard S Lazarus. *Emotion and adaptation.* Oxford University Press, 1991.

[136] Nicole Lazzeri, Daniele Mazzei, Lorenzo Cominelli, Antonio Cisternino, and Danilo Emilio De Rossi. Designing the mind of a social robot. *Applied Sciences*, 8(2):302, 2018.

[137] Colin Wayne Leach, Russell Spears, and Antony SR Manstead. Parsing (malicious) pleasures: Schadenfreude and gloating at others' adversity. *Frontiers in Psychology*, 6, 2015.

[138] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. Empathic robots for long-term interaction. *International Journal of Social Robotics*, 6(3):329–341, 2014.

[139] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.

[140] Heather C Lench and Zari Koebel Carpenter. What do emotions do for us? In *The Function of Emotions*, pages 1–7. Springer, 2018.

[141] Nangyeon Lim. Cultural differences in emotion: differences in emotional arousal level between the east and the west. *Integrative Medicine Research*, 5(2):105 – 109, 2016.

[142] Patricia L Lockwood and Miriam C Klein-Flügge. Computational modelling of social cognition and behaviour—a reinforcement learning primer. *Social cognitive and affective neuroscience*, 2020.

[143] Anibal Antonio Lopez, Paulo Renato Lourenço, Isabel Dimas, and Cludia Figueiredo. Pjawsn–escala portuguesa do bem-estar afectivo no trabalho: Contributos para a sua validação. *A emoção nas organizações*, pages 155–178, 2012.

[144] Juan M Machimbarrena, Esther Calvete, Liria Fernández-González, Aitor Álvarez-Bardón, Lourdes Álvarez-Fernández, and Joaquin González-Cabrera. Internet risks: An overview of victimization in cyberbullying, cyber dating abuse, sexting, online grooming and problem-

atic internet use. *International journal of environmental research and public health*, 15(11):2471, 2018.

[145] Alexandra Main and Carmen Kho. A relational framework for integrating the study of empathy in children and adults. *Emotion Review*, 12(4):280–290, 2020.

[146] Javier Marín-Morales, Juan Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Scilingo, Mariano Alcañiz, and Gaetano Valenza. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports*, 8(1), 2018.

[147] Stacy Marsella, Jonathan Gratch, Paolo Petta, et al. Computational models of emotion. *A Blueprint for Affective Computing*, 11(1):21–46, 2010.

[148] Stacy C Marsella and Jonathan Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.

[149] Abigail A Marsh. The neuroscience of empathy. *Current opinion in behavioral sciences*, 19:110–115, 2018.

[150] Joseph W Matiko, Stephen P Beeby, and John Tudor. Fuzzy logic based emotion classification. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4389–4393, 2014.

[151] David Matsumoto and Linda Juang. *Culture and Psychology*. Nelson Education, 2016.

[152] Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215, 1992.

[153] Albert Mehrabian. Analysis of the big-five personality factors in terms of the PAD temperament model. *Australian journal of Psychology*, 48(2):86–92, 1996.

[154] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996.

[155] Ben Meuleman and Klaus R Scherer. Nonlinear appraisal modeling: An application of machine learning to the study of emotion production. *IEEE Trans. on Affective Computing*, 4(4):398–411, 2013.

[156] Marina Micheli. Social networking sites and low-income teenagers: between opportunity and inequality. *Information, Communication & Society*, 19(5):565–581, 2016.

[157] Sylvia A Morelli, Matthew D Lieberman, and Jamil Zaki. The emerging study of positive empathy. *Social and Personality Psychology Compass*, 9(2):57–68, 2015.

[158] Juan E Moreno, Mauricio A Sanchez, Olivia Mendoza, Antonio Rodríguez-Díaz, Oscar Castillo, Patricia Melin, and Juan R Castro. Design of an interval type-2 fuzzy model with justifiable uncertainty. *Information Sciences*, 513:206–221, 2020.

[159] Mohammad Obaid, Ruth Aylett, Wolmet Barendregt, Christina Basedow, Lee J Corrigan, Lynne Hall, Aidan Jones, Arvid Kappas, Dennis Küster, Ana Paiva, et al. Endowing a robotic tutor with empathic qualities: Design and pilot evaluation. *International Journal of Humanoid Robotics*, 15(06):1850025, 2018.

[160] Patrick D O'Brien and Richard C Nicol. Fipa—towards a standard for software agents. *BT Technology Journal*, 16(3):51–59, 1998.

[161] Magalie Ochs, David Sadek, and Catherine Pelachaud. A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems*, 24(3):410–440, 2012.

[162] Kevin N Ochsner and Elizabeth Phelps. Emerging perspectives on emotion–cognition interactions. *Trends in Cognitive Sciences*, 11(8):317–318, 2007.

[163] Andrew Ortony, Gerald L Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge University Press, 1990.

[164] Mark D Packard and Thomas A Burnham. Do we understand each other? toward a simulated empathy theory for entrepreneurship. *Journal of Business Venturing*, 36(1):106076, 2021.

[165] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. Empathy in virtual agents and robots: a survey. *ACM Trans. on Interactive Intelligent Systems (TiiS)*, 7(3):11, 2017.

[166] Stefan Palan and Christian Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[167] Brian Parkinson, Peter Totterdell, Rob B Briner, and Shirley Reynolds. *Changing moods: The psychology of mood and mood regulation*. Longman London, 1996.

[168] Bert Peeters. *Semantic primes and universal grammar: Empirical evidence from the Romance languages*, volume 81. John Benjamins Publishing, 2006.

[169] Arthur Pewsey, Markus Neuhäuser, and Graeme D Ruxton. *Circular statistics in R*. Oxford University Press, 2013.

[170] Rosalind Wright Picard. *Affective Computing*. The MIT Press, 1997.

[171] Gordon D Plotkin. *A structural approach to operational semantics*. Aarhus university, 1981.

[172] Robert Plutchik. The nature of emotions. *American Scientist*, 89:344–350, 2001.

[173] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125, 2017.

[174] Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3):715–734, 2005.

[175] Stephanie D Preston. A perception-action model for empathy. *Empathy in mental illness*, 1:428–447, 2007.

[176] Stephanie D Preston and Frans BM De Waal. Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, 25(1):1–20, 2002.

[177] Priberam Dictionary. Priberam English–Swedish Dictionary. https://dicionario.priberam.org, 2021 (accessed September 17, 2021).

[178] Brian A Primack, Sabrina A Karim, Ariel Shensa, Nicholas Bowman, Jennifer Knight, and Jaime E Sidani. Positive and negative experiences on social media and perceived social isolation. *American Journal of Health Promotion*, 33(6):859–868, 2019.

[179] Anand S Rao. Agentspeak (l): Bdi agents speak out in a logical computable language. In *European workshop on modelling autonomous agents in a multi-agent world*, pages 42–55. Springer, 1996.

[180] Anand S Rao, Michael P Georgeff, et al. BDI agents: from theory to practice. In *1st Int. Conf. on Multiagent Systems*, volume 95, pages 312–319, 1995.

[181] Zeeshan Rasool, Naoki Masuyama, Md Nazrul Islam, and Chu Kiong Loo. Empathic interaction using the computational emotion model. In *2015 IEEE Symposium Series on Computational Intelligence*, pages 109–116. IEEE, 2015.

[182] Rainer Reisenzein, Eva Hudlicka, Mehdi Dastani, Jonathan Gratch, Koen Hindriks, Emiliano Lorini, and John-Jules Ch Meyer. Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange. *IEEE Trans. on Affective Computing*, 4(3):246–266, 2013.

[183] Ranier Reisenzein. Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, 67(3):525, 1994.

[184] Helen Riess. The science of empathy. *Journal of patient experience*, 4(2):74–77, 2017.

[185] Jaime A Rincon, Angelo Costa, Paulo Novais, Vicente Julián, and Carlos Carrascosa. A dynamic emotional model for agent societies. In *Int. Conf. on Practical Applications of Agents and Multi-Agent Systems*, pages 169–182. Springer, 2016.

[186] Jaime A Rincon, Angelo Costa, Paulo Novais, Vicente Julian, and Carlos Carrascosa. Using emotions in intelligent virtual environments: the ejacalive framework. *Wireless Communications and Mobile Computing*, 2017, 2017.

[187] Giacomo Rizzolatti, Luciano Fadiga, Vittorio Gallese, and Leonardo Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141, 1996.

[188] Sérgio Hortas Rodrigues, Samuel Mascarenhas, João Dias, and Ana Paiva. A process model of empathy for virtual agents. *Interacting with Computers*, 27(4):371–391, 2015.

[189] Luis-Felipe Rodríguez, J Octavio Gutierrez-Garcia, and Félix Ramos. Modeling the interaction of emotion and cognition in autonomous agents. *Biologically Inspired Cognitive Architectures*, 17:57–70, 2016.

[190] Ira J Roseman and Craig A Smith. Appraisal theory: Overview, assumptions, varieties, controversies. In *Appraisal Processes in Emotion: Theory, Methods, Research*, pages 3–19. Oxford University Press, 2001.

[191] Ramon Ruiz-Dolz, Stella Heras, José Alemany, and Ana García-Fornes. Towards an argumentation system for assisting users with privacy management in online social networks. In *CMNA@ PERSUASIVE*, pages 17–28, 2019.

[192] James A Russell. Evidence of convergent validity on the dimensions of affect. *Journal of personality and social psychology*, 36(10):1152, 1978.

[193] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[194] James A Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145, 2003.

[195] James A Russell and Lisa Feldman Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*, 76(5):805, 1999.

[196] James A Russell, Maria Lewicka, and Toomas Niit. A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology*, 57(5):848–856, 1989.

[197] Cheryl L Rusting and Randy J Larsen. Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models. *Personality and individual differences*, 22(5):607–612, 1997.

[198] Jelle Saldien, Kristof Goris, Bram Vanderborght, Johan Vanderfaeillie, and Dirk Lefeber. Expressing emotions with the social robot probo. *Int. Journal of Social Robotics*, 2(4):377–389, 2010.

[199] Miguel A Salichs and Maria Malfaz. A new approach to modeling emotions and their use on a decision-making system for artificial agents. *IEEE Trans. on Affective Computing*, 3(1):56–68, 2012.

[200] Klaus R Scherer. Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion*, 7(3-4):325–355, 1993.

[201] Klaus R Scherer, Angela Schorr, and Tom Johnstone. *Appraisal processes in emotion: Theory, Methods, Research*. Oxford University Press, 2001.

[202] Harold Schlosberg. The description of facial expressions in terms of two dimensions. *Journal of experimental psychology*, 44(4):229, 1952.

[203] Marc Schröder, Paolo Baggia, Felix Burkhardt, Catherine Pelachaud, Christian Peter, and Enrico Zovato. Emotionml–an upcoming standard for representing emotions and related states. In *Int. Conf. on Affective Computing and Intelligent Interaction*, pages 316–325. Springer, 2011.

[204] Frank Schweitzer, Tamas Krivachy, and David Garcia. An agent-based model of opinion polarization driven by emotions. *Complexity*, 2020, 2020.

[205] Elizabeth A Segal. Social empathy: A new paradigm to address poverty. *Journal of Poverty*, 11(3):65–81, 2007.

[206] Suzanne C Segerstrom and Gregory T Smith. Personality and coping: Individual differences in responses to emotion. *Annual review of psychology*, 70:651–671, 2019.

[207] Sofia Serholt and Wolmet Barendregt. Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th nordic conference on human-computer interaction*, pages 1–10, 2016.

[208] Nigel Shadbolt, Paul R Smart, JR Wilson, and S Sharples. Knowledge elicitation. *Evaluation of human work*, pages 163–200, 2015.

[209] Ali Sharifara, Mohd Shafry Mohd Rahim, and Yasaman Anisi. A general review of human face detection including a study of neural networks and haar feature-based cascade classifier in face detection. In *2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, pages 73–78. IEEE, 2014.

[210] Cristiana S Silva, Glívia AR Barbosa, Ismael S Silva, Tatiane S Silva, Fernando Mourão, and Flávio Coutinho. Privacy for children and teenagers on social networks from a usability perspective: a case study

on facebook. In *Proceedings of the 2017 ACM on Web Science Conference*, pages 63–71, 2017.

[211] Lennart Sjöberg, Elisabeth Engelberg, et al. Emotionell intelligens-teori och empiri i ett psykologiskt perspektiv. *Emotional intelligence-theory and empirical research in a psychological perspective*, 2003.

[212] Julia Stietz, Emanuel Jauk, Sören Krach, and Philipp Kanske. Dissociating empathy from perspective-taking: Evidence from intra-and inter-individual differences research. *Frontiers in psychiatry*, 10:126, 2019.

[213] Karsten Stueber. Empathy. *International Encyclopedia of Ethics*, 2013.

[214] Synonymer Dictionary. Synonymer Swedish dictionary of synonyms. https://www.synonymer.se, 2021 (accessed September 17, 2021).

[215] Joaquín Taverner, Bexy Alfonso, Emilio Vivancos, and Vicent J. Botti. Using different agent personalities to obtain different agent behaviours. In *Proc.gs of the Workshop on Affective Computing and Context Awareness in Ambient Intelligence (AfCAI 2018)*, 2018.

[216] Joaquin Taverner, Bexy Alfonso, Emilio Vivancos, and Vicente Botti. Modeling personality in the affective agent architecture GenIA3. In *Int. Conf. on Agents and Artificial Intelligence (ICAART)*, pages 236–243, 2018.

[217] Joaquin Taverner, Bexy Alfonso, Emilio Vivancos, and Vicente J Botti. Integrating expectations into Jason for appraisal in emotion modeling. In *IJCCI (ECTA)*, pages 231–238, 2016.

[218] Joaquin Taverner, Ramon Ruiz, Elena del Val, Carlos Diez, and Jose Alemany. Image analysis for privacy assessment in social networks. In *International Symposium on Distributed Computing and Artificial Intelligence*, pages 1–4. Springer, 2018.

[219] Joaquín Taverner, Emilio Vivancos, and Vicente Botti. Towards a computational approach to emotion elicitation in affective agents. In *International Conference on Agents and Artificial Intelligence*, pages 275–280, 2019.

[220] Joaquin Taverner, Emilio Vivancos, and Vicente Botti. A multidimensional culturally adapted representation of emotions for affective computational simulation and recognition. *IEEE Transactions on Affective Computing*, 2020.

[221] Joaquin Taverner, Emilio Vivancos, and Vicente Botti. A fuzzy appraisal model for affective agents adapted to cultural environments using the pleasure and arousal dimensions. *Information Sciences*, 546:74–86, 2021.

[222] Auke Tellegen, David Watson, and Lee Anna Clark. On the dimensional and hierarchical structure of affect. *Psychological science*, 10(4):297–303, 1999.

[223] Paul T Van Katwyk, Suzy Fox, Paul E Spector, and E Kevin Kelloway. Using the job-related affective well-being scale (jaws) to investigate affective responses to work stressors. *Journal of occupational health psychology*, 5(2):219, 2000.

[224] Renata Vieira, Álvaro F Moreira, Michael Wooldridge, and Rafael H Bordini. On the formal semantics of speech-act based communication in an agent-oriented programming language. *Journal of Artificial Intelligence Research*, 29:221–267, 2007.

[225] Douglas N Walton. *Practical reasoning*. University of California Press, 2020.

[226] Henry M Wellman. Theory of mind: The state of the art. *European Journal of Developmental Psychology*, 15(6):728–755, 2018.

[227] Anna Wierzbicka. Human emotions: universal or culture-specific? *American Anthropologist*, 88(3):584–594, 1986.

[228] Travis J Wiltshire, Emilio JC Lobato, Daniel S McConnell, and Stephen M Fiore. Prospects for direct social perception: a multi-theoretical integration to further the science of social cognition. *Frontiers in Human Neuroscience*, 8:1007, 2015.

[229] Lauren Wisp. History of the concept of empathy. *Empathy and its development*, pages 17–37, 1987.

[230] Lauren Wispé. *The psychology of sympathy.* Springer Science & Business Media, 1991.

[231] Joshua Wondra and Phoebe Ellsworth. An appraisal theory of empathy and other vicarious emotional experiences. *Psychological review*, 122(3):411, 2015.

[232] Markus Wrbouschek and Thomas Slunecko. Moods in transition: Theorizing the affective-dynamic constitution of situatedness. *New Ideas in Psychology*, 62:100857, 2021.

[233] Özge Nilay Yalçın. Evaluating empathy in artificial agents. *arXiv preprint arXiv:1908.05341*, 2019.

[234] Özge Nilay Yalçın. Empathy framework for embodied conversational agents. *Cognitive Systems Research*, 59:123–132, 2020.

[235] Özge Nilay Yalçın and Steve DiPaola. A computational model of empathy for interactive agents. *Biologically inspired cognitive architectures*, 26:20–25, 2018.

[236] Özge Nilay Yalçın and Steve DiPaola. Modeling empathy: building a link between affective and cognitive processes. *Artificial Intelligence Review*, pages 1–24, 2019.

[237] Michelle S M Yik, James A Russell, Chang-Kyu Ahn, José Miguel Fernández Dols, and Naoto Suzuki. Relating the five-factor model of personality to a circumplex model of affect: A five-language study. In *The five-factor model of personality across cultures*, pages 79–104. Springer, 2002.

[238] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[239] Lotfi Asker Zadeh. The concept of a linguistic variable and its application to approximate reasoning. In *Learning systems and intelligent robots*, pages 1–10. Springer, 1974.

[240] Tomasz Zaleskiewicz and Jakub Traczyk. Emotions and financial decision making. In *Psychological perspectives on financial decision making*, pages 107–133. Springer, 2020.

[241] Eduard Zell, Katja Zibrek, and Rachel McDonnell. Perception of virtual characters. In *ACM Siggraph 2019 Courses*, pages 1–17. ACM, 2019.

[242] Guozhen Zhao, Yan Ge, Biying Shen, Xingjie Wei, and Hao Wang. Emotion analysis for personality inference from EEG signals. *IEEE Trans. on Affective Computing*, 9(3):362–371, 2018.

[243] Ke Zhong, Tianwei Qiao, and Liqun Zhang. A study of emotional communication of emoticon based on russell's circumplex model of affect. In *International Conference on Human-Computer Interaction*, pages 577–596. Springer, 2019.