Modeling Empathy in Embodied Conversational Agents

by

Özge Nilay Yalçın

M.Sc., Middle East Technical University, 2014
B.Eng., Istanbul Technical University, 2010

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

© Özge Nilay Yalçın 2019
SIMON FRASER UNIVERSITY
Summer 2019

Copyright in this work rests with the author. Please ensure that any reproduction
or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name: **Özge Nilay Yalçın**

Degree: **Doctor of Philosophy**

Title: **Modeling Empathy in Embodied Conversational Agents**

Examining Committee:

Chair: Marek Hatala
Professor

Steve DiPaola
Senior Supervisor
Professor

Fred Popowich
Supervisor
Professor
School of Computing Science

Alissa N. Antle
Internal Examiner
Assistant Professor

Jonathan Gratch
External Examiner
Professor
Institute for Creative Technologies
University of Southern California

Date Defended: **August 28, 2019**

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Embodied conversational agents (ECAs) are designed with the goal of achieving natural and effortless interactions with humans by displaying the same communication channels we use in our daily interactions (e.g. gestures, gaze, facial expressions, verbal behaviors). With advances in computational power, these agents are increasingly equipped with social and emotional capabilities to improve interaction with the users. Recently, research efforts are focused on modeling empathy, which is a human trait that allows us to share and understand each other's feelings. The emerging field of computational empathy aims to equip artificial agents with empathic behavior, which has shown great promise in enhancing the human-agent interaction. However, two issues arise in this research endeavor.

Firstly, even though a variety of disciplines have extensively examined empathic behavior, there is minimal discussion on how that knowledge can be translated into computational empathy research. Second, modeling and implementing a complex behavior such as empathy poses a great challenge on fluent and automated integration of these behaviors to achieve real-time and multi-modal interaction with ECAs. This thesis aims to model and implement empathy in embodied conversational agents while focusing on both of these issues.

To achieve this goal, an extensive literature review of the definitions and models of empathy from various disciplines is provided. Building upon this background knowledge, a model of empathy is presented that is suitable for interactive virtual agents, which includes three hierarchical layers of behavioral capabilities: emotional communication competence, emotion regulation and cognitive mechanisms. This dissertation further provides suggestions on how to evaluate perceived empathy of such a system, as there are no agreed-upon standards or best-practices in this novel field on evaluation metrics. Following the establishment of these theoretical foundations, levels of empathic behavior were implemented into an ECA with real-time spoken conversation capabilities that include synchronized gestural and emotional behavior. Evaluations of this system, which is called M-PATH, showed that the proposed levels of behavioral capabilities resulted in an increase in the perception of empathy as well as the perceived usefulness, human-likeness and believability of the agent. This dissertation further demonstrates that implementing empathic behaviors in artificial agents would not only improve our interaction but can also enhance our understanding of empathy by providing us with a controlled environment to implement and test our theories.

Keywords: empathy; affective computing; embodied conversational agents; human-computer interaction

Dedication

... to me.

Acknowledgements

I would like to express my appreciation to the people that made this work possible by supporting me in various ways in the last four years.

First of all, I would like to thank my senior supervisor Prof. Steve DiPaola for adopting me at a point where I was thinking of giving up on my PhD adventure in Canada and restoring my faith in academia again. His vision and enthusiasm has always inspired me to develop my ideas and pushed me further than I thought I could go. Also my supervisor Prof. Fred Popowich for his guidance, encouragement, support, and his hard questions which made me think about my research from different perspectives.

I would also like to thank my friends and lab-mates at SIAT and iVizLab. Thank you Michael, Ulysses, Procheta, and our undergrads Ioana and Marie, for their help during the project.

And my partner in academic misery, Burcu. Your dark humor and friendship have been priceless. I hope life will throw us and Burçin into each other's direction much more often. My little ball of happiness, Aynur, I am glad that I decided to be roommates with a crazy person. Let's do more projects together and pretend we are mature.

Last but not least, my family. Anniş and babiş. I love you. Thank you for encouraging me in all the steps I took in my life. It was your unconditional trust that led me believe in myself and love that lifted my spirits up. Abiş, I am so lucky that my best friend and mentor is my brother. I miss your company and our laughs the most. And Sergey. The best thing about these four years was meeting you. You made everything about this journey amazingly better. I can't wait to discover more in life with you.

Finally, I would like to acknowledge the financial support. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2019-06767] and the Social Sciences and Humanities Research Council of Canada (SSHRC) [435-2017-0625]. Additional funding was provided by Ebco Eppich Graduate Scholarship in Intelligent Systems, School of Interactive Arts and Technology in Simon Fraser University (SFU). Additionally, the presentation of the papers included in this thesis in respective conferences were supported by the travel grants from SFU Graduate Student Society (GSS), School of Interactive Arts and Technology (SIAT) and Association of Computing Machinery (ACM).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Imagine the first time you read one of your favorite books, trying to put yourself into the shoes of the main character and picture what you would do if you were in their shoes. Or imagine standing in front of a painting and trying to grasp what is about it that makes you feel so sad. Or the time when you felt concerned about the people you have never met before, after watching the news about how their lives got destroyed by a natural disaster.

All of these scenarios have one crucial element in common: empathy. Empathy can be defined as the capacity to understand, experience and resonate to another person's emotional state (de Waal, 2007; Davis, 1994). Empathic behavior is an important part of our social relationships as a species. Levels of empathic behavior considered as an adaptation that enables us to understand each other (Coplan & Goldie, 2011) and act ethically (Decety & Cowell, 2014; A. Smith, 1959), which is also found in a variety of animals (de Waal & Preston, 2017). It is said to be the "glue that holds communities together" while allowing us to be in tune with others and caring for those in need (de Waal, 2010, p. X).

Since the introduction of the concept by the scholars of aesthetics theory in the beginning of the 20th century (Wisp, 1987), the interest in empathy has dramatically increased (Coplan & Goldie, 2011). The importance of empathy has been repeatedly stated in business applications (Goleman, 2005), medical practice (Halpern, 2001; Mercer & Reynolds, 2002; Hojat, 2007), politics (Crawford, 2014), education (Feshbach & Feshbach, 2009), animal studies (Bekoff, 2010; de Waal, 2010) and moral theory (A. Smith, 1959; Decety & Cowell, 2014). As a result, a variety of disciplines have contributed to defining, modeling and examining empathic behavior such as philosophy (A. Smith, 1959; M. Smith, 2011; Slote, 2007), psychology (Hoffman, 2001; Batson, 2009), psychiatry (Clark, 2014), neuroscience (Goldman, 2011) and ethology (de Waal & Preston, 2017).

In a world where the interactions are increasingly becoming digital, empathy has also caught the attention of researchers in human computer interaction (HCI). With recent advances in technology and artificial intelligence (AI) research, the idea of developing interactive agents capable of understanding and expressing socio-emotional behaviors has become increasingly popular. These agents can take on the role of educators, assistants and even companions give rise for the needs for interactive virtual agents to behave more sensitively to human emotions (Picard, 2014). The novel field of computational empathy emerges from this environment with the aim of equipping artificial agents with the ability to act emphatically (Paiva, Leite, Boukricha, & Wachsmuth, 2017).

This thesis is motivated by the notion of modeling empathic behavior in artificial agents as a way of enhancing our interaction with these systems and improving our understanding of empathy by being able to study it outside of human behavior. We have chosen embodied conversational agents (ECAs) as our implementation environment, as they are suitable candidates for simulating natural interaction between humans and AI systems.

1.1.1 Empathy in Embodied Conversational Agents

Embodied conversational agents (ECAs) are agents that can interact with users with a multi-modal, situated (and often anthropomorphic), and real-time interaction to emulate a similar experience of human-to-human conversational interaction (Cassell, Bickmore, Campbell, & Vilhjálmsson, 2000). ECAs are argued to provide a more effortless interaction due to the natural input and output modalities that are intuitive to humans. Figure 1.1 shows examples of ECAs that we created to be used in our system.

Figure 1.1: Examples of characters with different genders and characteristics that are created to be used in our system.

In this research, we chose our characters and animation system to allow for a high-level of naturalness and realism as a foundation for the empathic behaviors. Realism of an agent could show a wide-range of differences from full-motion photo-realistic quality 3D avatars to a plain text "screen name" on a chat window. Avatar realism can be characterized as possession of visual and/or behavioral attributes of a human being (Bailenson, Yee, Merget, & Schroeder, 2006). The characters we used in this dissertation was aimed at achieving a high-level visual realism that includes human-like facial and bodily features, as well as the behavioral realism that could be created from the natural movement of these features.

Two of the characters were created at the USC Institute of Technology that can be controlled via the Smartbody character animation system as realistic 3D models (Thiebaux, Marsella, Marshall, & Kallmann, 2008; Shapiro, 2011). We are currently creating additional characters in our research group under my supervision, as an extension of this dissertation follows a similar idea with varying levels of characteristics such as age, gender and race (also discussed in Section 9.3.4). As we seek a balance between behavioral realism and visual realism that would allow us to manipulate the behaviors of the character to express empathic behavior, we did not use the highly photo-realistic avatars that would not allow for proper synchronized behavior control (Feng, Rosenberg, & Shapiro, 2017) and more cartoon-like characters that would have high behavior control but do not equip the visual characteristics for expressing natural facial and bodily gestures (Bailenson et al., 2006).

It has been argued that verbal information is the primary source of understanding the context of affective information in empathic behavior (Omdahl, 1995). Moreover, the face-to-face conversation has been proposed to be the natural basis for the human-computer interaction, which is behaviorally rich and includes different modalities such as face and body gestures, posture changes, gaze, paralinguistic parameters, and linguistic context (Gratch et al., 2002). As humans tend to use their existing communicational and social skills while interacting with artificial agents, it is reasonable to embody the agents with natural communication capabilities to create a familiar basis to interact with. The use of input and output modalities that will correspond to these means such as cameras and visual feedback, haptic interaction techniques, microphones and sound are called perceptual interfaces (Oviatt & Cohen, 2000). These techniques allow users to apply natural interaction skills to interact with technology which in turn, create a rich, natural and efficient interaction with computers (Turk & Robertson, 2000). These notions suggest that conversational agents can be beneficial in empathy research, as they can be used to provide additional context to affective and social signals. In this work, we define our system as a multi-modal conversational framework that embodies multiple input and output modalities to produce and recognize conversational signals. However, equipping the system with these capabilities might have consequences on the usability of the system.

In order to successfully design such a computational system, the multi-modal input and output capabilities should be synchronously generated in real-time while keeping track of the context and the conversational functions. Timing, synchronization and context-relatedness of these behaviors would influence the perception of speaker's competence (Reeves & Nass, 1996; Churchill, Cook, Hodgson, Prevost, & Sullivan, 2001), the meaning conveyed by these signals (Vinciarelli et al., 2012), the quality (Jaimes & Sebe, 2005) and the naturalness of the interaction (Cassell, Vilhjálmsson, & Bickmore, 2004). Equipping a real-time system with such rich behavioral capabilities can be computationally demanding and may result in a decrease in the fluency of behavior. Most ECA frameworks attempt to ensure real-time and synchronous behavior in the agent while maintaining fluent high-level communication capabilities. However, this is still an ongoing research effort.

Real-time behavior in human feedback has been defined as feedback that is "provided in real-time, as the speaker is articulating (or having trouble articulating) her utterance. This means that the feedback mechanism can not wait until after the speaker has finished to calculate the feedback." (D. Traum, DeVault, Lee, Wang, & Marsella, 2012, p. 275). The main reason to aim for a real-time behavior in a computational system is "to reduce the cognitive load on users or to enable them to increase their productivity without the cognitive load on them increasing" (Turner, 1986 as cited in Laffey et al., 1988). In this research, we targeted real-time response to achieve natural, human-like and fluent behavior that would not cause the interaction to be disturbed by the rate of feedback. One important factor of real-time expectations when using interactive realistic characters is the synchronization lip movements during speech, as humans are very sensitive to misaligned lip-synchronization (Steinmetz, 1996). Another factor is synchronization of the gestures to speech of the character, as many studies suggest a tight relationship between phases of gesture and speech (Wagner, Malisz, & Kopp, 2014).

Consequently, equipping embodied conversational agents with empathic capabilities is a rewarding yet challenging endeavor. An empathic interaction requires successful recognition of emotions from the user, process it according to the mechanisms of empathy, and respond to the user by generating emotional and synchronized behavior in real-time. Simulating an empathic ECA is, therefore, a multidisciplinary effort which must capture the richness and dynamics of human behavior in conversation in addition to the complexity of the empathic behaviors. The motivation for this thesis is to examine the possibility of modeling empathy in embodied conversational agents and increase our understanding of empathic behavior within the context of human-computer interaction.

1.2 Research Overview

My research investigates the problem statement: How can we model empathy in embodied conversational agents? Empathy, as a complex socio-emotional phenomenon, requires theoretical conceptualization in order to be modeled in a virtual agent. A computational model and implementation of empathy are necessarily an interdisciplinary effort, due to their reliance on the well-established theories from various disciplines such as philosophy, psychology and neuroscience (Paiva et al., 2017). However, these existing models, theories and even definitions of empathy can dramatically differ within and between disciplines. Moreover, even though the research area of computational empathy is moving towards developing standard methods; the variations in the capabilities, embodiment, application areas and goals of the agents make it unreasonable to reach to a consensus. Additionally, modeling empathic behavior within a real-time and multi-modal conversational framework would pose further challenges to the fluent, natural and believable interaction.

Nevertheless, research efforts on modeling empathy based on theoretical models can provide valuable contributions to our understanding of empathy by allowing us to test our assumptions efficiently. The objective of this dissertation is therefore not to reach a singular solution but to provide guidelines and recommendations on how a theoretically-sound implementation of empathic behavior can be implemented in an interactive system, that can be backed up by empirical data.

Consequently, the research problem of this dissertation is addressed by conceptualizing this overarching objective into a series of research questions and developing a methodology to systematically focus on each question.

1.2.1 Research Questions

Building on the overarching goals and motivations mentioned above, this thesis focuses on the following research questions:

- **RQ1** What are the theoretical requirements/components for an empathic agent?

  This question lies the theoretical foundation of the empathy research in AI and is aimed to specify the components required for simulating empathy mechanisms in artificial agents. The question is aimed to explore the theoretical foundations of empathy and empathic behavior that can be applied to the computational empathy research. In order to achieve this, a comprehensive background on the theoretical foundations that are proposed by various fields and the existing techniques that are being used in the computational empathy research need to be reviewed.

- **RQ2** How can a computational empathy model be described to be efficiently used by a virtual agent?

  Following the theoretical background on empathy, this question is aimed to provide a theoretically grounded model of empathy for interactive agents. This step is required to allow for a translation between empathy models in the current literature and the computational model that can be used in implementing the empathic virtual agent.

- **RQ3** How can we evaluate an empathic virtual agent?

  Simulating empathy in artificial agents is a relatively new area of research, and the evaluation methods are not standardized. This research question is aimed to help define evaluation methods derived from psychology and HCI research.

- **RQ4** How can an empathy model be implemented in an embodied conversational agent framework?

The final research question in the thesis is aimed at incorporating the results of the previous questions to the implementation of the theoretical empathy components in an embodied conversational agent. To be able to achieve this, state-of-the art methods from related research disciplines such as human-computer interaction, affective and social computing as well as artificial intelligence will need to be examined in addition to generating new methods. The answer to this question might dramatically change according to the context of the interaction and agent capabilities. Therefore, it is important to limit the focus on a specific use-case scenario.

1.3 Contributions

Concerning the overarching goal of simulating empathic behavior in embodied conversational agents, this thesis will present the research efforts which are aimed to address the research questions formulated above. Contributions of this thesis can be categorized into three distinct processes: construction of the theoretical foundations, the implementation and the evaluation (see Figure 1.2).

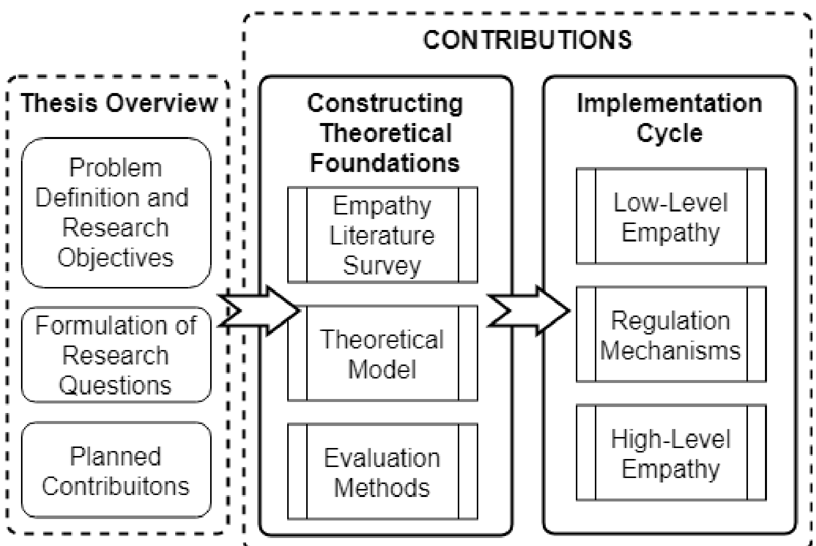

Figure 1.2: Figure shows the overview of the research presented in this thesis. Contributions of the thesis can be categorized into the construction of theoretical foundations and the implementation of the model within the implementation cycle.

Construction of the theoretical foundations is achieved in three steps that are covered in the following three chapters of this thesis. The first step of this process includes the construction of a theoretical model that is grounded in the interdisciplinary theoretical background of empathic behavior in humans. This step is intended to answer the **RQ1**, which is addressed in Chapter 2 as an extensive survey for empathy research. Informed by this theoretical background, the next step is to construct a model of empathy that is suitable for embodied conversational agents. This step is intended to answer **RQ2**, and it is addressed in Chapter 3. The last step in constructing the theoretical foundations is to gather suitable evaluation methods for an empathic embodied conversational agent. This step is aimed to address **RQ3** in Chapter 4.


Figure 1.3: The setup of an interaction between M-PATH and the user. The agent shows empathic listening behavior (left) and responds to the user input (right). User input is gathered from the web-cam and microphone that can be seen in the setup.

Following these theoretical foundations, the implementation process is initiated with the theoretical model and the hypothesized components required for simulating empathy in our embodied conversational agent: M-PATH. Figure 1.3 and Figure 1.4 shows an example interaction setup of our system. We used this setup to be minimally intrusive and close to human conversational interaction in a face-to-face interaction (Figure 1.3) or a video-conference setting (Figure 1.5). Both of these interaction settings include a microphone and a web-cam that can collect multi-modal input from the user, and a screen with audio output that can show the audio-visual behaviors of the agent. Face-to-face interaction setting was aimed to achieve a comfortable daily interaction, in a room with a reasonable lighting. The video-conference setup is similarly intended to mimic regular interaction through a video-conferencing scenario. These setups are designed to establish ecological validity, by approximating the materials and setting of the study to reflect the real-world scenario in question, namely the Hospital Intake Scenario (for the face-to-face interaction) and the video-chat scenario with a friend (video-conferencing interaction).

The modeling of empathy requires multiple components that are needed to be integrated into the system iteratively, which is formulated in the implementation cycle. Here, implementation of the system includes the perceptual module, behavior generation module and a behavior controller that allows for real-time natural conversational interaction with basic levels of empathic behavior (see Figure 1.6). This step is intended to address **RQ4** within Chapters 5, 6, 7 and 8, where the last two chapters also overlap with the last part of the contributions.

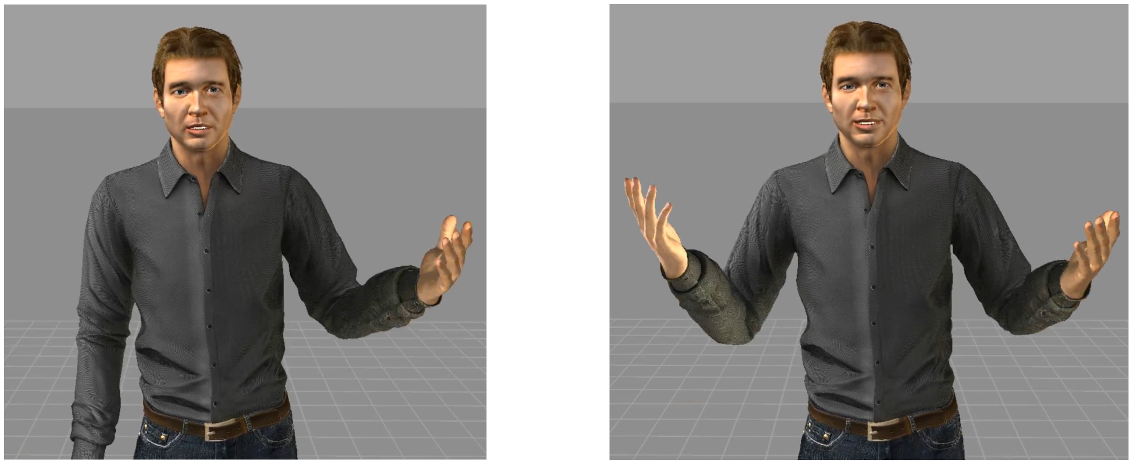

Figure 1.4: Examples of automated affective gesture generation behavior that accompanies the speech of our agent. Figures show fear (left) and happy (right) emotions with beat gestures.

The empathic behaviors and M-PATH's interactive behavior are evaluated in Chapters 7 and 8, which addresses **RQ3** and **RQ4**. As the first step of evaluation, we focus on the generation and perception of low-level empathic capabilities such as mimicry and affect matching along with the backchanneling behavior (Chapter 7). Figure 1.5 shows examples from the empathic listening scenario interaction video. Following this step, we provide the generation of empathic concern and consolation behavior, as a result of emotion regulation and coping mechanisms that result from emotion regulation behavior (Chapter 8). The implementation of the high-level cognitive mechanisms such as appraisal and perspective taking are outside the scope of this thesis. However, we further provide theoretical requirements and suggestions for future implementation of such a system (Chapter 3 and 5).

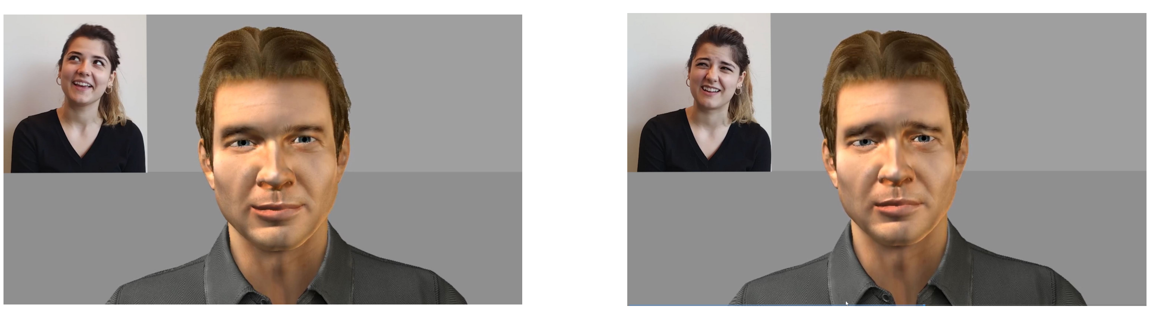

Figure 1.5: The setup of an interaction between M-PATH and the user, where the agent shows empathic listening behavior with affect matching or mimicry. Details of this study is mentioned in Chapter 7.

Each phase will have corresponding contributions to the implementation and evaluation of the components of the proposed empathy model. As it can be seen in Figure 1.2, the first design cycle will be the implementation of the communication competence, and the

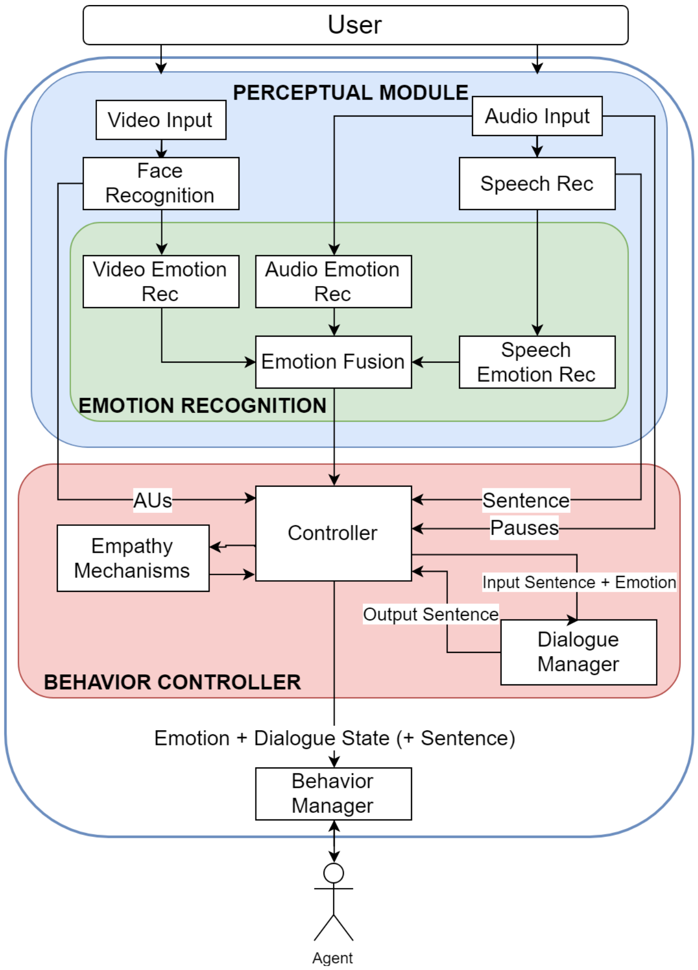

Figure 1.6: Overall framework of M-Path that shows the main modules.

second cycle will be the implementation of the emotion regulation that will be built on top of the previous implementation. After each design cycle, evaluations will be carried out to refine both the model and the system and results will be published. Table 1.1 shows the corresponding papers for each contribution step and chapter of this thesis.

The contributions of the thesis can be summarized as follows:

1. Determine the theoretical requirements for levels of empathic behavior to be translated into the computational empathy research

   - Provide a survey of the theoretical underpinnings of empathic behavior in humans and the implementations in the computational empathy research (**RQ1**)
   - Examine the evaluation methods for computational empathic agents (**RQ3**)
   - Describe potential computational model for an embodied conversational agent following the theoretical models (**RQ2**)

2. Provide the implementation of the computational model of empathy in an embodied conversational agent (**RQ4**). Our implementation is highly modular (see Figure 1.6) and includes several state-of-the-art methods that allow for flexibility in future modifications for incorporating new methods.

    - **Perceptual Module**: The perceptual capabilities of our agent includes real-time facial emotion and voice recognition via a web-cam and a microphone setup. This module includes facial, linguistic and voice-based emotion recognition components in addition to speech recognition functionality that allows fluent multi-modal conversational interaction. We use deep machine learning methods for emotion recognition such as: Haar Cascade Classifiers for face detection, facial component detection, and Convolutional Neural Networks for emotion classification for face and voice. For sentiment analysis, we integrated the state-of-the-art rule-based methods into our system. Similarly, we used cloud-based speech recognition methods to incorporate the most advanced methods within our framework.
    - **Behavior Generation Module**: This module automatically generates emotional facial and bodily gestures that are synchronized with the speech of the agent, allowing for listening, thinking and speaking behaviors during real-time conversational interaction. This module automatically controls the facial and bodily gesture generation according to the dynamics of emotion that are synchronized with the speech of the agent. We used a combination of rule-based and filtering systems to achieve a natural balance between fluidity and expressivity of the behaviors. We used BML standard for controlling synchronized verbal and non-verbal behavior for Smartbody Behavior Controller system.
    - **Behavior Controller**: This central component acts as a decision maker that includes a dialogue manager and an empathy mechanisms component. The Dialogue Manager allows for a real-time natural conversation flow with flexibility in turn-taking. We use TF-IDF based similarity models along with fine-tuning the pre-trained language representation models for an efficient dialogue interaction by using neural networks. The Dialogue Manager also includes a survey controller system that assures a goal-directed behavior to conduct surveys in the database, using a memory-based reasoning system. The Empathy Mechanisms component provides decisions on emotional behavior with respect to the selected empathy level, mood and personality of the agent.

3. Evaluate levels of empathic behavior in an embodied conversational agent, following the evaluation methods that are examined (**RQ3, RQ4**). This is done by designing, implementing and statistically analyzing our system through a series of empirical studies aimed to conduct feature-based and system-based evaluations of the behaviors of our system.

- **Low-Level Empathic Behaviors** : In addition to the baseline conversational backchanneling behavior, mimicry and affect matching behaviors are evaluated with the system as low-level empathic capabilities.
- **Empathic Concern and Consolidation** : We implement and evaluate basic emotion regulation capabilities to provide empathic concern and consolidation behavior during a goal-directed conversation scenario.
- **Cognitive Mechanisms** : We provide details on the requirements to achieve cognitive mechanisms in an embodied conversational agent and suggestions for future implementation.

1.4 Thesis Structure

This cumulative thesis is structured as a series of nine chapters, each with a discrete focus. This chapter provides introductory information, and Chapter 9 provides a summary and discussion of conclusions. The remaining seven chapters are written as stand alone, peer reviewed papers that are aimed to address each of the specific research questions of the thesis. As such, there is inevitably overlap of information between chapters. Chapters 2, 3 and 4 provides the theoretical background, model and evaluations methods in order to implement empathy in interactive agents. Chapters 5, 6, 7 and 8 focuses on the implementation and methodological evaluation of levels of empathic behavior in an embodied conversational agent based on these theoretical foundations.

Chapter 2 provides a comprehensive background on the theoretical foundations of empathy research from a variety of disciplines in order to systematically examine the definitions, components and models of empathy that can be used in computational empathy research. Following this survey, this chapter provides a detailed examination and discussion of the current techniques used to model empathy in interactive agents. Further, this chapter lays down the foundational theoretical understanding to determine the requirements for modeling and implementation of empathy in embodied conversational agents.

Chapter 3 proposes a model of empathy for interactive agents, following the theoretical foundations presented in the previous chapter. The proposed model is inspired by the influential evolutionary model by Preston and de Waal (de Waal, 2007; Preston & de Waal, 2002), and focuses on the hierarchical organization of the components of empathy. These components include emotional communication competence, emotion regulation mechanisms and cognitive mechanisms. The chapter further provides possible ways of implementing these components, while giving examples on affective and social computing research.

Chapter 4 provides a systematic approach to evaluate empathy in artificial interactive agents, as the novel field of computational empathy lacks standards and specific methods. This chapter focuses on the well-established evaluation methods from empathy research in

Table 1.1: Main Contributions of the Thesis

CHAPTER	CONTRIBUTIONS	PUBLICATION
1. Introduction	- Motivation - Research Questions	-
2. Modeling Empathy: Building a Link Between Affective and Cognitive Processes	- Theoretical Background on Empathy - Empathy Models and Computational Empathy Review - Collecting Requirements for an Empathic Agent - Overview of Challenges and Future Directions	AI Review Journal
3. A Computational Model of Empathy for Interactive Agents	- Theoretical Model of Empathy - Components and Implementation Requirements	BICA Journal
4. Evaluating Empathy in Artificial Agents	- Review Empathy Evaluation in Humans - Suggest Evaluation Methods for Empathy in Artificial Agents - Draw Parallels from HCI research to Incorporate Variables that can Effect the Evaluation	ACII 2019 Conference
5. Empathic Listener Framework for Embodied Conversational Agents	- Define a Framework that Incorporates the Theoretical Model from Chapter 3 - Explain the Implementation Steps for the Modules of the Framework	Cognitive Systems Research Journal
6. Automated Affective Gesture Generation for Embodied Conversational Agents	- Implementation Details for Emotion Regulation Components of the Embodied Conversational Agent	Journal on Multimodal User Interfaces
7. Levels of Emotional Contagion in an Embodied Conversational Agent: Implementation and Evaluation	- Implementation Details for Low-Level Empathic Behavior - Evaluation of the System	COGSCI 2019 Conference
8. M-Path: A Conversational System for the Empathic Virtual Agent	- A Dialogue System for the Empathic Conversational Agent - Providing Implementation Details and Evaluation	BICA 2019 Proceedings
9. Conclusion	- Discussion of the Findings - Future Directions	-

humans and HCI research to translate these methods into computational empathy research. This chapter is crucial both in providing guidelines on how an implementation of interactive empathic agent can be evaluated and in initiating the dialogue on the development of standards for a variety of applications and systems.

Chapter 5 presents the computational framework and the implementation of the proposed theoretical model of empathy in an embodied conversational agent. It provides the details of implementation for communication competence and affect regulation component for low and mid-level empathic behavior. This chapter also proposes the requirements for future implementation of the cognitive mechanisms that are required for higher level empathic behavior such as appraisal mechanisms, user modeling and perspective taking.

Chapter 6 shows the details of the implementation of emotion regulation mechanisms, building on top of the empathy framework described in the previous chapter. Implementation includes parameters for automated emotional gesture generation for facial and bodily gestures that accompany the speech of the agent. Emotion regulation processes that are linked to long and short term dynamics of emotions such as mood and personality are generated as a result of this implementation.

Chapter 7 demonstrates the results of the implementation and evaluation of the low-level empathic behavior in an embodied conversational agent. Using the implementation of the agent described in the previous chapters, this paper examines the effect of mimicry and affect matching behaviors of the agent during the conversation. The paper focuses on the evaluation of these behaviors using the metrics based on Chapter 4. The results of three consecutive studies in this chapter revealed that both mimicry and affective matching behaviors were perceived as more empathic than the baseline listening behavior, where the difference between these behaviors was only significant when the agent verbally responded to complex emotional behaviors.

Chapter 8 provides the details of the dialogue management system within the framework of our conversational empathic agent. The goal-driven dialogue system is used in a Psychological Counseling Service intake scenario, where our agent conducts surveys to students in a text-based chat environment. The empathic response generation capabilities of the agent are evaluated in a preliminary study, that showed significant results on the perceived empathy of the agent. The evaluation metrics used were based on Chapter 4. Results of our studies show that our implementation of emotional and coping responses as empathic behavior in our conversational agent leads to an increase in the perception of empathy, usefulness as well as human-likeness and believability of the agent.

Chapters 2, 3, 7 and 8 have already been published. Chapters 4 and 5 are accepted and in publication in peer reviewed journals and conferences. Chapter 6 is submitted and awaiting reviewers' decisions for publication. I am the primary author of all the chapters and the sole author of Chapters 4 and 5. The scope of the contributions is described at the beginning of each section. The references for all the chapters were added collectively in the References section of the thesis. Additional material that are published within the scope of this research that are not included in the main body of this thesis are included in the appendices. The code generated for this thesis and the research data are also shared as links that can be found in Appendix A and Appendix B, respectively.

Chapter 2

Modeling Empathy: Building a Link Between Affective and Cognitive Processes

This chapter is published in AI Review Journal as: Yalçın, Ö. N. & DiPaola, S. (2019). **Modeling Empathy: Building a Link Between Affective and Cognitive Processes**. *Artificial Intelligence Review*, pp. 1-24. Springer. doi:10.1007/s10462-019-09753-0.

Contributions: I was the main contributor to this paper. I was responsible for surveying the empathy literature, investigation and conceptualization of the main concepts of this paper. I was also responsible for writing the original draft. Prof. DiPaola was supervising and reviewing the paper.

Any attempt of modeling a complex phenomenon initially requires a theoretical grounding of the concept. Following this idea, the first step I took was to survey the vast literature of empathy from various disciplines such as philosophy, psychology, psychiatry and neuroscience. I mainly focused on the definitions, behaviors and models of empathy that were assigned by these disciplines to be able to create a guideline and establish theoretical foundations for a computational model of empathy that is suitable for embodied conversational agents. Following this necessary step of gathering theoretical components of empathic behavior, I provided a detailed review of the current techniques in artificial empathy research in interactive agents. I examined each technique in the light of the theoretical models and definitions, in addition to the details of implementation.

2.1 Abstract

Computational modeling of empathy has recently become an increasingly popular way of studying human relations. It provides a way to increase our understanding of the link between affective and cognitive processes and enhance our interaction with artificial agents. However, the variety of fields contributing to empathy research has resulted in isolated approaches to modeling empathy, and this has led to various definitions of empathy and an

absence of common ground regarding underlying empathic processes. Although this diversity may be useful in that it allows for an in-depth examination of various processes linked to empathy, it also may not yet provide a coherent theoretical picture of empathy. We argue that a clear theoretical positioning is required for collective progress. The aim of this article is, therefore, to call for a holistic and multilayered view of a model of empathy, taken from the rich background research from various disciplines. To achieve this, we present a comprehensive background on the theoretical foundations, followed by the working definitions, components, and models of empathy that are proposed by various fields. Following this introduction, we provide a detailed review of the existing techniques used in AI research to model empathy in interactive agents, focusing on the strengths and weaknesses of each approach. We conclude with a discussion of future directions in this emerging field.

2.2 Introduction

Recent developments in technology and artificial intelligence (AI) research has created an environment that allowed virtual agents to become a part of our daily lives, not only as mere tools, but as assistants and even companions. The increasing capabilities of computational systems for sensing and processing have made it possible to process social and affective cues and naturally interact with people. Computational empathy is a novel paradigm that emerges from this environment. It refers to the ability of computational systems to understand and respond to the thoughts, feelings, and emotions of humans.

Empathy is a complex socio-emotional phenomenon that can be defined as the capacity to understand and react towards the feelings, thoughts, and experiences of others. This capacity allows us to perceive another's point of view by resonating with their emotions. It has been argued that empathy plays an essential role in forming and maintaining social interactions; it helps to coordinate actions, understand the intentions of others, and facilitate prosocial behavior between individuals (Omdahl, 1995). Since empathy is so important to social interactions, the integration of empathic capability for computational systems would also be useful. It could enhance interactive systems such as educational applications, medical assistants, companions, psychotherapy, and gaming applications where social capabilities are of great importance. Moreover, as an amalgamation of the affective and cognitive processes, computational empathy can provide us grounds to examine the link between emotions and cognition.

Since the introduction of the term "empathy" in the 19th century, a diversity of disciplines have contributed to its study. These sources of knowledge are invaluable for the computational models of empathy. However, the rich variety of fields contributing to the empathy research has resulted in a number of approaches used in computational modeling of empathy. This in turn has led to vague definitions and the absence of common ground in conceptualizing the processes underlying empathy. This paper presents a holistic approach to empathy modeling that integrates the theoretical and empirical background of empathy research from various fields. Our goal is to provide a systematic summary of the field to aid theoretical positioning and further research in the emerging field of computational empathy.

In order to achieve this, the following section provides a comprehensive theoretical background of empathy to capture the variety of behaviors attributed to empathy from this broad spectrum of fields (Section 2.3). This section is followed by computational models based on three main components that can be used to methodologically compare these approaches: emotional communication competence, emotion regulation and cognitive mechanisms (Section 2.4). The subsequent section presents the methods used in implementing the computational models of empathy in interactive agents as theory-driven and data-driven approaches (Section 2.5). We conclude with a discussion of these approaches (Section 2.6) and future directions (Section 2.7) in this promising field.

2.3 Theoretical Background of Empathy

Empathy has been an influential concept in ethics and moral theory (Hoffman, 2000; Slote, 2007; A. Smith, 1959), aesthetics theory (M. Smith, 2011), social/developmental psychology (Hoffman, 2000; Batson, 2012), clinical psychology (Clark, 2014), and neuroscience (Goldman, 2011; de Waal, 2010), which all have followed the foundational work in philosophy. Contributions from this rich variety of fields has resulted in a number of definitions, functions and proposed components of empathy, which pose a challenge to studying empathy (Coplan, 2011).

However, research efforts are starting to converge towards certain aspects of empathic behavior that merge emotional and cognitive phenomena. A definition of empathy as an umbrella term for any type of process triggered by observing the others' emotional states and activating one's own (de Waal & Preston, 2017), has been used as an attempt to cover this broad range of behaviors. Behaviors that have been used to define empathy (Batson, 2009; Coplan & Goldie, 2011; de Waal & Preston, 2017; Leiberg & Anders, 2006) can be listed as: motor mirroring (mimicry), affective matching (emotional contagion), concern about another's state (sympathy, empathic concern), consolidation behavior (altruistic helping), understanding another's state and thoughts (theory theory), imagining another's thoughts (perspective-taking), and projection of self into another's situation (projective empathy, simulation theory). Empathy research often focuses on each of these phenomena individually. However, empathy research in psychology, neuroscience, and ethology suggests these behaviors represent the levels of empathic behavior that are connected through evolutionary processes, where each layer is built on top of the other, without replacing the layer before (de Waal & Preston, 2017).

Most AI research on empathy has depicted a binary categorization and evaluation of empathy as empathic and non-empathic behavior (Prendinger & Ishizuka, 2005; Brave, Nass, & Hutchinson, 2005; Rodrigues, Mascarenhas, Dias, & Paiva, 2015). More recently, researchers (Ochs, Sadek, & Pelachaud, 2012; Boukricha, Wachsmuth, Carminati, & Knoeferle, 2013) have represented levels of empathy by modulating emotions using the personality, mood, and social link parameters together with an appraisal of the situation. Another line of research in robotics focuses on developmental aspects of empathy (Asada, 2015; Lim & Okuno, 2015), which suggests that empathic behavior is learned to some extent and need to be based on neurological foundations.

Despite much progress, these research efforts are still in their infancy, and they highlight the need for developing and testing the models of empathy to challenge and solidify existing approaches. A computational model of empathy should support the theoretical background and the empirical research gathered from related fields. In the later sections, we provide an overview of definitions and models of empathy, including recent findings from psychology, neuroscience and ethology research.

2.3.1 Definitions of Empathy

One of the earliest mentions of empathic behavior can be seen in the work of Hume (1739), who used the term "sympathy" as a notion that is related to what we now refer to as low-level or affective empathy (see (Wisp, 1987) and (Nowak, 2011) for a complete history of the term). Hume conceptualized sympathy as an automatic process that allows for emotion contagion, morality and aesthetic pleasure while focusing on the communicability of affect for further cognitive processing. Similarly, Adam A. Smith (1959) distinguished sympathy from both pity and compassion by assigning it a communicative function. He refers to sympathy as a higher-level process and cognitive capacity that is related to perspective taking and simulation of other minds. In these early definitions of sympathy Hume and Smith focus on different levels of empathy that continue to be widely adopted today.

The first mention of the term "empathy" is made in the late 19th century (as cited in (Coplan & Goldie, 2011; Wisp, 1987)) as "Einfühlung" by Theodor Lipps, which soon became a fundamental concept for aesthetics and understanding other minds. The fundamental work of Lipps inspired a generation of thinkers to study "Einfühlung" (which means "feeling into") in several domains, including psychology, aesthetics, and philosophy; Titchener later translated this term to English as "empathy" (Titchener, 1909).

The term "empathy" later came to largely replace "sympathy" in a number of fields, and the concept acquired additional behavioral functions. In addition to being an involuntary and affective feedback mechanism, it was assigned new functions as observations about others' affective, behavioral and mental states (Kohut, 2011). It came to refer to an affective response that reflects another's situation rather than one's own, (Hoffman, 2000), and an imagination of another's thought processes (Stueber, 2006).

As empathy has continued to be studied by psychology, neuroscience and artificial intelligence, the number of related processes and cognitive capacities involved in empathic behavior have increased. (Coplan, 2011) lists these empathic capabilities as emotional contagion, caring behavior, perspective taking, being affected by emotions of others, theory of mind, as well as a combination of these. Some researchers focus on a narrow definition of empathy, specifying that empathy only occurs when the observer is in an affective state as a result of having imagined or observed the target's affective state with a clear self-other distinction (de Vignemont & Singer, 2006). This definition excludes from empathy concepts such as sympathy, emotional contagion, and personal distress. Consistent with this narrow definition, Coplan (2011) argues that self-other differentiation is at the core of empathy and it can only be said to exist when three features of empathy are present: affective matching, other-oriented perspective taking, and self-oriented perspective taking.

In contrast to these narrow views of empathy, new research by several researchers has focused on a broader definition of empathy that aims to generate empathy models that can explain how the interaction of various processes may give rise to the broad spectrum of concepts. In the following section, we will examine these models of empathy and demonstrate how they are supported by psychology, neuroscience and ethology research.

2.3.2 Models of Empathy

There are two main approaches to theoretical models of empathy: categorical and dimensional. Categorical models focus on the two distinct levels of empathy mechanisms that are referred to as high/cognitive and low/affective empathy. In contrast, dimensional models propose a more multidimensional system where the levels of empathy are functionally integrated.

The former categorical approaches identify two levels of the term 'empathy' by distinguishing affective from cognitive levels of empathy. These different levels are also referred to as low/high empathy (Goldman, 2006), basic/re-enactive empathy (Stueber, 2006) and mirroring/constructive empathy (Goldman, 2011). (Omdahl, 1995) attempts to connect these levels by categorizing empathy as affective, cognitive and a mixture of the two. Similarly, Hoffman (2000) argues for five modes of empathic arousal that include low-level processes of mimicry and afferent feedback, classical conditioning, and association to one's own experience; as well as high level mediated association and perspective-taking, where a mixture of these modes can be observed in an individual.

Affective (or low-level) empathy is considered to be the automatic mimicking of another's emotional response as one's own. This level is suggested to arise from the Perception-Action mechanism (PAM), which has its biological roots in mirror neurons (Preston & de Waal, 2002). Some researchers distinguish the mirroring of bodily states in a category other than cognitive and emotional empathy, calling it kinesthetic empathy (Wood, 2016; Hagendoorn, 2004). Cognitive (or high-level) empathy is defined as the ability to understand the target's mental state by imagining how they feel. This ability is related to perspective taking and theory of mind (ToM) (Batson, 2009).

Evidence for both affective (low-level) and cognitive (high-level) empathy has been found in various mammalian species, suggesting a Darwinian assumption on underlying processes (de Waal & Preston, 2017; Preston & de Waal, 2002). Recent findings in neuroscience, especially the discovery of mirror neurons (Rizzolatti & Fabbri-Destro, 2010), also support an evolutionary basis of empathy. Research on mirror neurons inspired empathy research in neuroscience and established a foundation for the functional architecture of empathy mechanisms in humans (Iacoboni, 2011; Goldman, 2011). From mimicry of motor actions to sharing affective experiences, the neural activation of self-experience of emotions while being exposed to another's emotional experience has been suggested to be the core mechanism for empathy (Singer & Lamm, 2009).

Evidence for cognitive (higher-level) empathic behavior comes from pathological studies, autism studies, and developmental psychology (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003). Research has found that a malfunction in high- or low-level empathy mechanisms can lead to a spectrum of social behavior disorders such as autism and sociopathy (Preston & de Waal, 2002). For example, cognitive empathy is found to be impaired in individuals with autism spectrum disorder, even though they may have an intact lower-level empathic function (Baron-Cohen, Tager-Flusberg, & Lombardo, 2013). People with autism are therefore able to understand the type of emotions a person expresses, without being able to understand the reasons for those emotions. Psychopathy, on the other hand, is related to an intact ability for perspective-taking but an inability to share the resulting emotions (Preston & de Waal, 2002). Following these observations and evidence from neuroscience, distinct neural routes were hypothesized as mirroring and reconstructive routes of empathy (Goldman, 2011).

On the other hand, dimensional approaches to modeling empathy consider it to result from a set of interrelated constructs. Researchers that advocate this approach point to the evidence from evolutionary and neurological mechanisms that affective and cognitive levels of empathy are interconnected (de Waal & Preston, 2017; Goldman, 2011). For example, de Waal (2010) argues for an evolutionary foundation for the emergence of empathy based on the evidence of different levels of empathy mechanisms in social animals other than humans.

One example of a dimensional model is from (Davis, 1994), who also stresses the importance of the evolutionary roots of empathy mechanisms. In this model, Davis proposes to separate the processes taken place within the observer and the outcomes of these processes as affective and cognitive outputs. The model focuses on studying an empathic "episode" where the observer is exposed to emotional stimuli from a target. This exposure will result in an empathic reaction, which can have different properties according to four related constructs: antecedents, processes, intrapersonal outcomes and interpersonal outcomes. Antecedents refer to the characteristics of the person (biological capacities, learning history) and the situation (strength of the observed emotion, the degree of similarity). Three mechanisms primarily constitute the processes that produce empathic behavior: non-cognitive mechanisms (mimicry, primary reaction); simple cognitive mechanisms (classical conditioning, labelling, direct association); and advanced cognitive mechanisms (linguistic associations, perspective taking). The intrapersonal (affective and non-affective) and interpersonal (relationship-related outcomes) outcomes refer to the empathic behaviors arising from these mechanisms. Most computational models (McQuiggan & Lester, 2007; Ochs et al., 2012) refer to a sub-set of intrapersonal outcomes called parallel and reactive outcomes. These consist of the affective responses that differ from the target's emotional behavior (reactive empathy) and resonate to the target's emotion (parallel empathy).

Following this view and the evidence of empathic behavior in social animals, De Waal and Preston proposed the Russian-doll model of empathy (Preston & de Waal, 2002; de Waal, 2007). In this model, (Preston & de Waal, 2002) proposes that the capacity of empathy arises from the simple emotional contagion mechanisms that are based in mirroring and Perception-Action Mechanisms in mammals (de Waal, 2007; de Waal & Preston, 2017). Starting from this basic involuntary mechanism as the foundational layer, the model consists of three hierarchical layers where each layer depends on the previous layers. The second layer requires emotional self-regulation mechanisms that give rise to empathic concern and consolidation behavior. The last layer consists of cognitive functions such as perspective taking and theory of mind. This model allows for a wide range of empathic behaviors and emotional patterns starting from simple reflex-like copying mechanisms to higher-level cognitive functions. This theoretical model has been widely used in computational empathy research (Yalçın & DiPaola, 2018; Asada, 2015).

Another significant model by (de Vignemont & Singer, 2006) proposes a contextual approach to modeling empathy where empathy is modulated by appraisal processes. In their model, the authors distinguish between low-level affective behaviors, such as emotional contagion, sympathy and personal distress (which they refer as "narrow empathy"), to the empathic behavior that involves higher-level appraisal processes. The appraisal processes are then further categorized into early and late appraisals, where the former involves the direct appraisal of the emotional stimuli and the latter includes modulation factors after the appraisal having taken place. This extra processing in the late-appraisals results in a slower response. According to this model, the modulation factors can be related to the features of emotions (valence, intensity, saliency), to the relationship between target and observer (affective link, familiarity, similarity), to context (appraisal of the situation), and to features of the observer (mood, personality, gender, age, emotional regulation capacities). Although this approach seems to disregard lower-level/affective empathy behaviors, the authors argue that these behaviors can still arise when the appraisal processes modulate the reflex-like processes such as mimicry.

Table 2.1 shows an overview of the most prominent theoretical models of empathy mentioned in this section. Although each model seems to have unique details, there is significant overlap between these models that can be systematically analyzed. In the following section,

Table 2.1: Theoretical Models of Empathy

Author	Emotional Communication	Emotion Regulation	Cognitive Mechanisms	Empathy Levels
(de Waal, 2007)	PAM mechanisms; motor mimicry; emotional contagion	Self-regulation	Perspective taking	Connected levels; degrees of empathy
Davis (1983, 1994)	Affective Empathy Parallel Outcomes Noncognitve Processes	Antecedents	Cognitive Empathy Cognitive Processes	Interrelated Categories
Omdahl (1995)	Affective Empathy	Emotion regulation	Cognitive Empathy	Interrelated categories
Hoffman (2000)	Mimicry Afferent feedback Classical conditioning	Empathic bias	Association to self-experience Mediated association Perspective Taking	Distinct categories
Stueber (2006)	Basic Empathy	-	Re-enactive Empathy	Interrelated Categories
Goldman (2006, 2011)	Low-Level Empathy Mirroring	-	High-level Empathy Constructive Empathy	Interrelated Categories
Coplan and Goldie (2011)	Affective Matching	Inhibitory and regulatory mechanisms	Self-other differentiation self and other-oriented perspective taking	All-or-none
De Vignemont and Singer (2006)	Emotional Capacities	Modulatory factors	Appraisal Processes	Intensity Levels

we will examine the similarities between these models to be used in the computational modeling of empathy.

2.3.3 A Systematic Categorization of Empathy Models

The various theoretical approaches to modeling empathy can be united as a set of cognitive and behavioral capacities, which we call components. In an attempt to arrive at a comprehensive computational model of empathy, we propose to classify these components as emotional communication, emotion regulation, and cognitive mechanisms. Others have suggested the similar categories of empathic responses (Paiva et al., 2017; Boukricha et al., 2013), empathy modulation and empathy mechanisms, in which low-level perceptual mechanisms are categorized together with high-level cognitive functions in the empathy mechanisms category. Our approach differs by separating low- and high- level functions and this is intended to reflect the distinct –but functionally integrated– routes of empathy responsible for various empathic behaviors.

The proposed components have been selected to reflect the broad spectrum of behaviors assigned to empathy as a result of different types of mechanisms. Each component has its roots in empathy and emotion research, which we will review in the subsections below. Assigning each empathic behavior to its required mechanisms eases the translation of the theoretical knowledge into the design and implementation of computational models.

Computational models can improve our understanding of empathy mechanisms as well as help enhance interactive agents by equipping those agents with socio-emotional capabilities (S. Marsella & Gratch, 2014). A computational model of empathy is therefore beneficial to the research community as it can provide a means for testing theoretical work. Regardless of the chosen theoretical model, a computational framework of empathy should reflect the broad spectrum of empathic behaviors that arise from the interaction of affective and cognitive processes. For a systematical comparison of the theoretical and computational approaches, we examine the models according to three main components: emotional communication competence, emotion regulation, and cognitive mechanisms.

Emotional Communication Competence Scherer (2010a) argues that any type of emotional behavior requires emotional production and recognition competence. This distinct capacity of perceiving, accessing, and generating emotions in order to influence behavior, reasoning, or understanding has been called as emotional intelligence in psychology (Salovey & Mayer, 1990). By definition, all levels and intensities of empathic behavior require the underlying mechanism of perceiving and expressing emotions. This core component is represented as Perception-Action Mechanisms (de Waal, 2007) in the Russian Doll Model of empathy, as mentioned previously. Neurological disorders caused by a dysfunction in this mechanism such as autism spectrum disorders and psychopathy suggest that empathic behavior on various levels are linked to this capacity (Preston & de Waal, 2002).

This communicational competency of the individual is also thought to affect the intensity of empathic responses (de Vignemont & Singer, 2006). Therefore, emotional communication competence is an essential component of empathy, which includes emotion recognition, expression, and representation. Low-level empathy, including variations of mirroring behavior such as motor mimicry, yawn contagion, and emotional matching, is a natural consequence of the basic interaction of the subcomponents of emotional communication competence.

Emotion Regulation This component is related to a range of social, psychological, and biological factors. It is argued that the low-level communicational mechanisms are shaped by regulatory processes such as the biological and psychological features of the observer and the relationship between target and observer (Davis, 1994; Paiva et al., 2017). The features of the observer (self-related parameters) may include unconscious characteristics that work on different timescales such as attention, arousal, mood, and personality (de Vignemont & Singer, 2006). Relationship-related parameters include social links such as familiarity, closeness, relatedness and perceived similarity (de Waal & Preston, 2017; de Vignemont & Singer, 2006; Hoffman, 2000). As these mechanisms are unconscious and automatic, they require inhibitory mechanisms to mediate consolation behavior (de Waal & Preston, 2017), which is a mid-level empathic behavior that is present in various mammals. Moreover, Coplan (2011) argues that inhibition of these mechanisms is crucial for higher-level empathic behavior such as other-oriented perspective taking.

Cognitive Mechanisms These capabilities include the appraisal and re-appraisal of the situation, theory of mind and mental simulation. Higher-level empathy mechanisms such as perspective taking and targeted helping require a conscious evaluation of the event and control over the low-level components. Appraisal theories of emotion focus on the connection between emotions and the goals, needs and desires of an individual, and propose that this connection may provide a foundation for high-level empathy mechanisms (Omdahl, 1995). There are many appraisal theories, but the underlying common assumption is that emotions are a result of subjective evaluation of events according to the goals and needs of the individual (Roseman & Smith, 2001). In stimulus check theory, Scherer (2001) proposes that appraisals can happen at various levels of cognitive processing. This idea also aligns with the Russian Doll model (de Waal, 2007), in which the highest most sophisticated component of empathy, the outer components, are fundamentally linked to the perceptual and regulatory inner layers. Moreover, appraisal processes can be followed by a re-appraisal of the situation (Lazarus, 1966). Understanding the effect of the event over the observed behavior, and simulation of the possible effects of the same event over the observer requires these appraisal and re-appraisal processes. Even though appraisal processes lay at the center of understanding of the cause and effect relationship between the event and observed emotion, higher-level empathic capabilities, such as self- and other-related perspective taking requires more than appraisal and re-appraisal processes. The ability to differentiate between self and other, and assigning distinct mental attributes to other minds is necessary for perspective-taking behavior. This ability is called the theory of mind (Premack & Woodruff, 1978), which is related to simulation theory and theory theory (Leiberg & Anders, 2006). However, theory of mind alone is not sufficient for perspective-taking without the ability to evaluate the event-action coupling. Hence, theory of mind and appraisal mechanisms should work together.

Table 2.2: Summary of empathy components, the mechanisms that are responsible and the corresponding behavior.

Empathy Component	**Mechanisms**	**Empathic Behavior**
Communication Competence	Emotion Recognition Emotion Expression Emotion Representation	Mirroring Affective Matching
Emotion Regulation	Features of the observer Features of the relationship	Empathic Concern Consolation
Cognitive Mechanisms	Emotional Appraisal Theory of Mind	Altruistic Helping Perspective-Taking

Table 2.2 shows an overview of the variety of behaviors linked to these components of empathy. The following section (Section 2.4) will examine the theoretical and methodological differences in computational models of empathy while focusing on these components: communication competence, emotion regulation, and cognitive mechanisms. Following this section, we will give further detail on the processes used to implement these models based on theory-driven (top-down) or data-driven (bottom-up) approaches (Section 2.5).

2.4 Computational Models of Empathy

Computational approaches of empathy in artificial agents can follow a variety of models, even though they might share the same theoretical foundations. These differences are usually due to the differences in the aim and context of the application. A virtual agent for teaching children, a healthcare robot for people with disabilities or a VR environment that is designed for assisting meditation would have a diverse set of behavioral capabilities and goals to show empathy. Additionally, some researchers might focus on creating a computational framework of empathy, whereas others may investigate the effect of empathic behavior on human-computer interaction.

This section examines approaches that are used or can be used for modeling empathy in artificial agents. In line with the theoretical background, we review current computational models of empathy in social agents focusing on the proposed categorization of empathy on three levels: communication competence, emotion regulation and cognitive mechanisms. We also refer to the relevant research on affective computing and social computing communities that address similar problems, which can be integrated into artificial empathy research.

2.4.1 Communication Competence

Affective computing research has focused on emotion recognition (Zeng, Pantic, Roisman, & Huang, 2009; D'mello & Graesser, 2012; Poria, Cambria, Bajpai, & Hussain, 2017) and expression in artificial agents for many decades (Calvo & D'Mello, 2010; Picard, 2014; Cambria, 2016). These advancements also tackle the problem of understanding multi-modal emotional content (Poria, Cambria, Howard, Huang, & Hussain, 2016), which is crucial for designing an interactive model that is capable of emotion recognition. However, the systems used in artificial empathy research are usually limited in their emotion recognition and expression capabilities.

An example of implementation (Yalçın, in press; Yalçın & DiPaola, 2020) and detailed examination of emotional communication competence can be seen in the recent work of (Yalçın & DiPaola, 2019). In this work, an embodied conversational agent performs empathic listening behaviors by employing three different mechanisms: backchanneling, mimicry and affective matching. The agent perceives the speech signal as well as a combination of face recognition parameters (Ekman and Friesen's Facial Action Units and facial landmark positions) and processes this information according to the selected empathic behavior. The system uses both categorical and dimensional scale to represent emotions in order to choose proper emotional expression feedback with facial expressions and head nods.

One of the most advanced affective communication techniques can be seen in the EMMA framework (Boukricha et al., 2013), which was integrated into an embodied virtual agent platform. The agent, MAX, uses Pleasure-Arousal-Dominance (PAD) space for expressing and representing emotions. Its expressive repertoire consists of Ekman and Friesen's Facial Action Units (AUs), emotional speech (based on prosody changes), and eye blinking and breathing frequencies. EMMA uses AUs to perceive the emotion of the interaction partner.

Prendinger and Ishizuka (2005) present the Empathic Companion in a job interview task where the emotional state of the user is mapped onto a valence-arousal dimensional space. The emotional states of the users are recognized using physiological data (skin conductivity and electromyography). The character can respond via text according to three different scenarios consisting of irritation, boredom and high arousal/high valence state. The agent does not have a high-level empathy model but rather can sympathize according to predetermined categories.

Rodrigues and colleagues (2015) only used facial expressions to recognize and express emotions. These expressions are coupled with text that either compliments or insults the other agents in the simulation environment. The emotions are presented as a tuple of “type”, “valence”, “intensity” and “cause” parameters based on OCC theory of emotions (Ortony, Clore, & Collins, 1990).

In the CARE framework McQuiggan and colleagues (2008) uses self-reported affective states of the users in ten available emotion categories. Using self-reporting in emotion recognition is controversial (Scherer, 2005; Calvo & D'Mello, 2010) and may be misleading. Also, the categorical approach in emotion classification disregards the intensity of the emotions (Russell, 1980; Scherer, 2005) and ignores blended emotions where human emotions are usually not isolated (Jaimes & Sebe, 2005). For emotion expression, the authors used predefined emotional sentences that were presented as text.

Brave et al. (2005) use a Blackjack game scenario where the agent can display emotions with picture-based facial expressions and textual expressions. These expressions were only classified in two categories: self-related and other related emotions. The authors did not capture the emotional expression of the users, but the empathic feedback was only given according to the situation of the user in the game. Another categorical approach can be seen in the iCat robot by Leite and colleagues (2014), who used the valence of nonverbal behavior as emotion recognition components. The expressive behaviors of iCat include spoken utterances that are divided into supportive categories of information support, tangible assistance, esteem support and emotional support.

Moridis and Economides (2012) use parallel empathy and reactive empathy based on Davis's (1994) definition of empathy. In their implementation of an empathic tutoring agent, they use six basic emotions to show pedagogical feedback to students' happy, sad and fear emotions extracted from facial expressions. The parallel empathy shows affective state matching, where reactive empathy is an emotional reaction to the user's emotions.

2.4.2 Emotion Regulation

Empathy regulation involves several factors that influence the extent of empathic behavior such as valence, intensity, and saliency of emotions, social relations, context, as well as mood, personality, gender, age and emotional repertoire of the agent (Paiva et al., 2017). Most of the regulation factors mentioned in the previous sections have been extensively studied in social computing research. Social computing research on personality, social link, and mood provides means of regulating the emotions based on the agent's characteristics. Some researchers (Ochs et al., 2012; Boukricha et al., 2013) used these regulation factors as a way of demonstrating different levels of empathy, where others only used a binary empathic-nonempathic classification (Brave et al., 2005; Prendinger & Ishizuka, 2005; Rodrigues et al., 2015).

Boukricha and colleagues (2013) used the distance between the mood of the empathizer and the modulated empathic emotion presented by the empathizer to measure the degree of empathy. They used empathy modulation factors such as mood, desirability for self, liking and familiarity. Liking and familiarity values ranged in [0, 1] scale and predefined modulation factors. It was not clear how the desirability-for-self parameter, which is introduced in the

OCC model of emotion, was calculated within the framework. The only dynamic parameter seems to be the mood of the agent, which changes with each interaction.

McQuiggan and colleagues (2008) created the CARE framework, where the agent learns to show empathy in parallel to the target's emotions or reactively to a specific set of attributes collected from the target. The authors prefer to use the self-reported emotion categorization of the user and train their model by using age, gender, context, empathic index and goal directedness features of the user. They used Naive Bayes, decision trees and SVM approaches to model empathic reactions.

Rodrigues and colleagues (2015) use affective link, similarity, mood and personality as modulation factors of the potential empathic emotions. The first two factors are parameters of the social relationship between the target and the empathizer, where the latter two are psychological factors that only concern the empathizer. Similarity is calculated by the distance between the intensities and valences of the emotions of empathizer and the target. In this sense, similarity is only calculated according to the emotional responses towards the event. Affective link and personality of the virtual agents are predetermined parameters.

Asada (2015) proposes a different approach for emotion regulation that does not mention any of the psychological, social or cultural parameters previously mentioned in this section. He proposes that emotion regulation is a part of the cognitive regulatory mechanisms that are intentional. His framework does not include the effect of mood, personality and similar factors.

2.4.3 Cognitive Mechanisms

Empathy mechanisms are capabilities to successfully understand the affective states of others. This also includes understanding the context, which is the situation the agent is in, that requires reasoning capabilities. de Vignemont and Singer (2006) proposed that the empathy process is tightly linked with appraisal dynamics and dependent on the observer's situation. Cognitive appraisal theories (Roseman & Smith, 2001) state that the subjective assessment of an event triggers emotions. The lack of cognitive appraisal and high-level reasoning capacities in empathy models can only account for lower-level or affective empathy. However, the higher level empathic processes such as other-oriented perspective taking require a clear distinction of self-other and theory of mind components.

Affective computing and emotion research, and especially research on appraisal theory is closely related to empathy theories (Omdahl, 1995). Affect is the result of the cognitive assessment of an individual where the situation and events are appraised. Computational models of empathy often use the appraisal models in emotion and affective computing research such as the OCC model (Ortony et al., 1990) or Scherer's appraisal model (2010b). Alternatively, other models of appraisal processes such as EMA (S. C. Marsella & Gratch, 2009), which is commonly used in embodied conversational agents, can also be used in these

Table 2.3: Theory-Driven Approaches to Empathy and Related Empathy Components

Author	Model	Empathy Components: Communication Competence	Empathy Components: Regulation	Empathy Components: Cognitive Mechanisms
Brave et. al. (2005)	Self and other oriented emotions	Emotional pictures text	gender	-
Prendinger & Ishizuka (2005)	Sympathy	Physiological responses text	-	-
McQuiggan et al. (2008)	Parallel and Reactive Empathy (Davis, 1994)	Self-reported affect recognition Emotional sentence construction	Age, Gender	Goal directedness
Moridis & Economides (2012)	Parallel and Reactive Empathy (Davis, 1994)	Facial emotion recognition Facial expression and text	-	-
Ochs et al. (2012)	Formal model of empathic emotions (Scherer, 2010b)	Dialogue with facial expressions	-	Appraisal model (Scherer, 2010b)
Lisetti et al. (2013)	Affective Empathy	Facial emotion recognition Verbal & nonverbal expression	User model	OCC Appraisal Theory (Ortony et al., 1990)
Boukricha et al. (2013)	Late Appraisal Model De Vignemont & Singer (2006)	Facial emotion recognition Facial expression Speech Prosody Blinking, Breathing	Mood desirability liking familiarity	Belief-Desire- Intention (Bratman, 1987) (Rao & Georgeff, 1991)
Rodrigues et al. (2015)	Narrow Empathy De Vignemont & Singer (2006)	Facial expressions text	Affective link Similarity, Mood Personality	OCC Appraisal Theory (Ortony et al., 1990)
Leite et al. (2014)	-	Facial expressions gaze, speech expressive behaviors	Mood	Appraisal model (Scherer, 2001)
Asada (2015)	Russian Doll Model (de Waal, 2007)	Facial expressions	-	Perspective- Taking Theory of Mind
Yalcin and DiPaola (2018)	Russian Doll Model (de Waal, 2007)	Emotion Recognition Emotion Expression Emotion Representation	Self and Other Related Modulation	Appraisal Reappraisal Perspective-taking

frameworks. However, appraisal theories alone cannot account for achieving the empathic perspective taking and self-other distinction.

Bates and colleagues (1994) suggests the role of emotion and appraisal theories in modeling believable agents. Omdahl (1995) states that most appraisals are communicated through verbal information, and therefore linguistic information is more critical in providing context than nonverbal behaviors. Speech act theory (Searle, 1969) can be used to translate appraisal theories into a dialogue context (Ochs et al., 2012). Ochs and colleagues (2012) propose a formal model of empathic emotions based on Scherer's appraisal model of emotion (2010b). Here, empathy is formulated as an emotion towards the user as a result of the successful or unsuccessful completion of the user goals.

In the work of Rodrigues and colleagues (2015), the appraisal system consists of the recognition of emotion according to the agent's own appraisals as well as the selection of the response and the modulation of that response using affective links, similarity, mood and personality. An important distinction here is that the agents use their own belief systems and goals to determine the appraisal, not the target's. This disregards the link between the Theory of Mind and empathy research, where the other's world view is needed to infer their beliefs and intentions. However, the potential response is selected in a way that allows mimicking of the same emotional response where it does not have a direct mapping of appraisal situations.

In the EMMA framework, Boukricha and colleagues (2013) use the Belief-Desire-Intention (BDI) model (Lefimann, Kopp, & Wachsmuth, 2006) in their version of the late appraisal model of de Vignemont and Singer (2006). Beliefs represent the agent's knowledge about the world, where desires represent the goals of the agent. Intentions are the action plans to achieve the goals. The BDI Framework (Lefimann et al., 2006) was initially used for action planning in embodied conversational agents and is suitable for the appraisal mechanisms of an empathy model.

User modeling and adaptation research also indirectly focuses on empathy mechanisms in a way where the agent adapts to the user's behavior over time. These mechanisms relate to the Theory of Mind (TOM), where a model of the interaction partner is made that allows the self and other distinction. Paiva and colleagues (2017) highlight the importance of context, history and user modeling in achieving high-level empathic behavior. Lisetti and colleagues (2013) adopt user modeling techniques for their embodied conversational agent to reason with the outcomes of empathic actions it takes. Their approach shows an interesting direction towards combining the recent techniques in social computing and use it to provide reasoning capacities that are required in higher-level empathic behavior.

2.5 Modeling Empathy in Artificial Agents

This section is a continuation of the previous section, which outlined the current approaches to computational models of empathy in terms of empathy categories. This section continues the survey on these theoretical approaches of modeling computational empathy and further focuses on how theoretical models are implemented in artificial agents in detail. We will adopt two main methods for implementing a computational model of empathy: theory-driven approach and data-driven approach.

Theory-driven approaches (also called top-down or analytical approaches) are models that are formulated based on theories and concepts of empathy. This approach is intended to provide an overview of the mechanisms related to empathy and then use these mechanisms to model the smaller components of the system hierarchically. This approach has a strong explanatory power where it allows for a basis to test the theoretical components systematically.

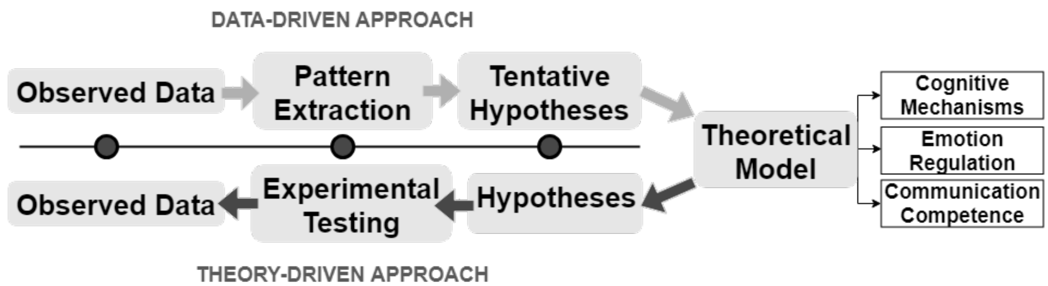

Figure 2.1: Top-down and bottom-up approaches for implementing empathy models in artificial agents.

On the other hand, data-driven approaches (also called bottom-up or empirical approaches) builds up from the empathic behavior data to train computational models and algorithms in an automated fashion. These models can be used for predicting or re-generating empathic behavior based on the desired application.

Hybrid approaches can also be used where theoretical components of empathy mechanisms are modeled separately to be used by the data-driven training methods. Given the success of emotion recognition algorithms (Cambria, 2016) and increasing computational power, these hybrid methods show great potential for the future of empathy research.

Both approaches can be used to reach or confirm a theoretical model, and have their strength and weaknesses. In the following sub-sections, we will review the computational approaches to empathy modeling while focusing on these methodological categorizations. Following this section, we will talk about the challenges of each method in the discussion (Section 2.6).

2.5.1 Theory-Driven Approaches

Theory-driven approach to modeling empathy has been increasingly popular within the artificial intelligence community due to the richness of theoretical models from a variety of disciplines, which we covered in Section 2.3.2. Researchers tend to use the components of these models as a starting point for their computational implementation. Depending on the selected model and details of each component, these models and implementations may differ in terms of the types of empathic behavior they cover.

A recent example of these approaches is seen in the component model of Yalçın and DiPaola (2018). This proposed model of empathy for interactive agents is highly inspired by the Russian Doll Model of Empathy (de Waal, 2007). In this model, the authors follow an evolutionary approach that connects behavioral patterns with related mechanisms. This model is composed of distinct but inter-connected components that are responsible for the various degrees of empathic behavior: emotional communication competence, emotion regulation and cognitive mechanisms. According to this model, the input from the event/context and the emotion are processed through the three layers of components. Each layer has information about the previous layer, as well as the processed information on these inputs. This approach allows for the possible implementation of low-level empathic behavior in isolation while providing a framework for simulating high-level empathic behaviors.

In his work, Asada (2015) also adopts the Russian Doll Model of empathy by de Waal (2007) from a developmental perspective. Although de Waal's model focuses on the evolutionary roots of the empathy mechanism rather than developmental changes, Asada's model draws parallels between the developmental theories of self-other distinction by proposing that developmental empathy should be a part of Cognitive Developmental Robotics (CDR). A similar approach that follows CDR is seen in the work of Lim and Okuno (2015), where empathic behaviors are learned through interaction of brain regions associated with specific mechanisms of empathy. However, their model only focuses on the affective/low-level empathic mirroring mechanisms.

Rodrigues and colleagues (2015) follow another approach that is linked with the theoretical approach of de Vignemont and Singer (2006) that states that humans do not feel empathy with every emotion and situation they encounter but instead, they select the response according to their appraisals. According to this view, empathy is never passive and is modulated according to similarity, affective link, mood and personality. Their proposed model is aimed to achieve more engaging and believable interactions with virtual humans.

Similarly, the virtual human EMMA (Boukricha et al., 2013) follows the late appraisal model of de Vignemont and Singer (2006). The EMMA framework consists of three modules: Empathy Mechanisms, Empathy Modulation and Expression of Empathy. A different approach by Ochs and colleagues (2012) provides a formal model of empathic emotions based on the theoretical formulations of Scherer (2010b) which were considered by Omdahl (1995) as the best candidate to model empathic emotions.

In their work, Ochs and colleagues (2012) represent emotions based on type (satisfaction, frustration, irritation, sadness and anger), intensity, emotion target, trigger event and the intention affected by the event. The empathic emotions use an additional variable for the target emotions in order to provide a logical formulation of them. One drawback of this approach is that it provides a very narrow view of empathy that fails to account for low-level empathic processes such as affective matching and high-level empathic processes such as perspective taking.

In their robotic empathic companion iCat, Leite and colleagues (2014) use predetermined empathic strategies that are inspired by the research on empathic behavior in teaching (Cooper, Brna, & Martins, 2000) combined with appraisal theories by Scherer (2010b) rather than following a clear theoretical framework. Similarly, some researchers focused only on mimicking (Gonsior et al., 2011; Riek, Paul, & Robinson, 2010) and affective matching (Lisetti et al., 2013; Hegel, Spexard, Wrede, Horstmann, & Vogt, 2006) capabilities, rather than presenting the entire spectrum of empathic behavior. Affective matching and mimicking behaviors have also been studied outside of empathy research as an indication of emotional intelligence (Burleson & Picard, 2007).

Additionally, some empathic behavior can be explained with current models of emotion that do not specifically address empathy. Scherer's sequential check model (2001) defines the appraisal processes as a sequence of evaluations done at different layers of processing: sensory-motor level, schematic level and the conceptual level. The sensory-motor level involves the innate assessment of the situation based on biological needs, such as hunger or pain. The schematic level is based on a learned assessment of cause-effect relationships. The conceptual level is based on symbolic cortical mechanisms that require consciousness. This framework can allow for different helping behavior which requires understanding causal relations of a stress-inducing event of another. Consolidation behavior, learned helping and targeted helping may result from the appraisal of the observed situation in different levels of processing. However, this model does not account for low-level mirroring and state matching behaviors, as they do not require an assessment of the situation. Mechanisms for theory of mind are also missing from this model, as it only accounts for appraisal of the situation in relation to self.

Broekens and colleagues (2008) suggest a formalism of the structure of appraisal. They provide a set of functions to perception, appraisal and mediation processes, as the core mechanisms of emotional behavior. Perception processes provide a link between the external world and internal representations. Appraisal processes use these current representations from working memory and assign appraisal values based on the evaluation. Lastly, mediating processes map these appraisals to emotion-component intensities, which is a set of emotional behaviors. This formalized model has a very similar underlying structure to the component model of Yalçın and DiPaola (2018). However, they differ in some significant ways. Firstly, similar to Scherer's theory, this is a serial model that does not allow any behavioral output before it follows the process line. Being a formalization of appraisal processes, this approach does not include the direct mapping between perception-action mechanisms which is responsible for mirroring and affective matching behavior. Moreover, this model does not allow for other-oriented perspective taking, which requires mechanisms for the theory of mind.

2.5.2 Data-Driven Approaches

Data-driven approaches to modeling of empathy are used to recognize, predict or generate empathic behavior from several types of behavioral data. The research on data-driven empathy can be said to be in its infancy, where there are only a few attempts to empirically model empathy using this approach in the last decade. The lack of datasets and differences in the labeling methods poses a challenge in building data-driven models of empathy (see Section 2.6).

Earlier work focuses on classifying empathy levels using linear models with various behavioral features. Xiao and colleagues (2014) build a computational model to classify therapist empathy levels by using prosodic features from the speech signal. They extracted pitch, energy, jitter, shimmer and utterance duration from the audio recordings of the psychotherapy interaction. These audio recordings were evaluated by three people using Motivational Interviewing Treatment Integrity (MITI) system (Moyers, Rowell, Manuel, Ernst, & Houck, 2016) to classify the empathy levels in seven categories. Similarly, Gibson and colleagues (2015) use motivational interviews to predict therapist empathy which was evaluated using the MITI Scale. They extract 13 psycholinguistic norm features from the text such as affective norms (Valence, Arousal, Dominance, Pleasantness), familiarity, gender ladenness, context availability in addition to n-grams.

Recent advances in machine learning allow vast amounts of data to be trained efficiently, where more complex models using neural networks can be used to link data to predict and even generate empathic behavior. Rashkin and colleagues (2018) provide a dataset of approximately 25k dialogues with empathic listener responses, called EmpatheticDialogues dataset. The authors used this dataset to fine-tune their pre-trained dialogue model in order to generate empathic responses to the utterances. The produced responses were evaluated by its performance on showing an "understanding of the feelings of the other person" using crowd-sourcing. Authors showed that their model produces more empathic responses both in retrieval and generation tasks.

Kumano and colleagues (2015) use video recordings of group interactions to provide a computational model to predict empathy. They use a Bayesian approach that connects the pairwise synchronization of gaze and facial expression information to the perceived empathy of the users. This is the first study that uses group dynamics in order to evaluate the empathy of the individuals as the input of the empathy model. Authors suggest that the collective evaluation of the individuals would remove the individual bias in scoring by allowing inter-group comparison. This idea, while being similar to the third-person evaluation of crowd-sourcing, provides a collective second-person evaluation. However, the evaluation is based on selecting the empathy levels on a 5-level empathy scale between “Strong Empathy” and “Strong Counter-Empathy”, where the definition of empathy was not provided.

Most of these data-driven models of empathy use a single modality, such as speech, text or video in order to train their models. However, the current advances in multi-modal emotion recognition techniques are most likely to change this trend soon. A recent distinct example is the OMG-Empathy Challenge (Barros, in press), which provided a dataset of interaction videos between a “speaker” and “listener”, where the speaker provides scripted stories. The interaction is recorded, and the felt empathy of the listener is then rated by the listener while watching this recording. One model from this challenge uses multi-modal inputs as well as the idea of synchronization in order to predict the empathy level of from the interaction video (Tan, Goel, Nguyen, & Ong, 2018).

Another notable approach can be seen in the hybrid model by McQuiggan and colleagues (2008), who developed a framework, CARE, where the empathy model trained on human-agent social interaction data that is capable of extracting intentions, actions, age, gender as well as affective states and biofeedback response. Learning from the user behavior in a simulation environment called Crystal Island, the authors aim to model artificial agents that can generate empathic responses. Their framework uses the theoretical model from Davis (1994) that is trained on user interaction data. Authors use Interpersonal Reactivity Index (Davis, 1983) to measure the empathic nature of the users, and their goal orientation is used to train the model using Naïve Bayes. This approach applies both the top-down and bottom-up methodology, which shows great promise for the future of empathy modeling research.

2.6 Discussion

Table 2.3 shows a summary of some of the empathy research regarding the methodology, theoretical background, and its coverage of three empathy components: emotional communication competence, emotion regulation, and cognitive mechanisms. Regarding the empathy components, only a few researchers used a complete spectrum of empathy mechanisms. However, research efforts are advancing in terms of applying a complete model of empathy and including a broader range of behaviors.

As the research on empathy in artificial agents is an emergent field, current models and techniques often fail to capture the broad spectrum of empathic behavior. There seems to be a tendency in artificial empathy research to refer to any system that can respond to affective signals as empathic, which ignores the cognitive and high-level processes involved in empathy mechanisms. Although it is not necessary for every attempt of modeling empathic behavior to adopt all the components of empathy, it is important that the theory, models, and implementation researchers provide are congruent with each other.

Research on social agents and affective computing has examined concepts that are similar to empathy without necessarily mentioning empathy itself. For example, communicational competence of agents has been studied in affective computing, especially in the areas of emotion recognition (Soleymani et al., 2017) and expression (Calvo & D'Mello, 2010). Also, user modeling and personality research in the field of social computing has investigated emotion regulation mechanisms as well as modeling self/other distinction, which are both key to high-level empathy. As they are not directly intended to model empathy, these research efforts only cover some aspects of empathy rather than providing a complete picture. However, they nonetheless provide valuable insights that are useful for modeling empathy in artificial agents.

2.6.1 Methodological Issues

Some of the challenges in computational modeling of empathy are specific to the chosen methodology. Theory-driven approaches allow for the translation of the observed relationships to a framework. These approaches are easier to test and have more explanatory power. However, theories are often more difficult to translate into computational implementation. This mismatch can result in ambiguity, which in turn can generate different implementations of the same theoretical model.

Additionally, even though theory-driven approaches are the only way to test the theories of empathy, a complete implementation and evaluation of each component is required before evaluating the models as a whole. Moreover, the implementation from vague theoretical definitions can be challenging, especially in higher-level cognitive mechanisms. It is not clear, for example, how to achieve theory-of-mind in a computational setting.

On the other hand, one of the main problems of the data-driven models is data collection. Most of the data-driven approaches focus on the behavioral cues of empathizer. However, being a complex socio-emotional phenomenon, directly linking visual and textual information to the evaluation of empathy can be misleading. The behaviors can only be considered as "empathic" when they are a response to another's emotional stimuli. Also, the annotation of the data, types of available modalities, and variety in context are all factors that affect the empathic behavior during an interaction.

2.6.2 Evaluation of the Model

Developing a reliable, sensitive, and valid measure of empathy for artificial agents is not an easy task. Evaluating agents on their empathic competence mostly relies on subjective user perception of a spectrum of characteristics, rather than on the application of objective measurement. When such evaluation tools are used, they tend to show many differences in the preferences of research subjects; subject preferences are very dependent on the capabilities of the computational model. There might also be differences between users in their definition of empathy due to its varying use in daily life. Considering the variations in the definitions and behavioral attributes, this vagueness in the term “empathy” poses an extra challenge to the research community.

To overcome this challenge, some researchers have chosen to measure perceived empathy by referring to empathy as “feeling with” (Boukricha et al., 2013), “felt sorry” (Rodrigues et al., 2015), “matching emotion” (McQuiggan et al., 2008), and “caring” (Brave et al., 2005). It is crucial to note which of empathic behavior these terms relate to during the evaluation of the system. Moreover, the assessment of the empathic behavior is often treated as a binary classification and coded as either an empathic or a non-empathic response. These approaches disregard the different levels of empathic responses as well as other components of empathy mentioned in the previous sections. In order to clarify the aims of the particular type of research, it is crucial to indicate which sense the term “empathy” is being used and to provide a precise positioning in the theoretical framework.

In psychology research, empathy of individuals is generally measured with the Empathy Quotient (Baron-Cohen & Wheelwright, 2004) which is a self-report scale that has been validated (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). Empathy Quotient and Systemizing Quotient are used to determine patients with Autism Spectrum Disorders, where the average scores of both autistic men and women have been found to be lower than their healthy counterparts (Baron-Cohen et al., 2003). Additional tests of empathy focus on distinct features of empathic capacity, such as perception of emotions (Tavassoli, Hoekstra, & Baron-Cohen, 2014) in pictures (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), voice (Golan, Baron-Cohen, Hill, & Rutherford, 2007) and movies (Golan, Baron-Cohen, Hill, & Golan, 2006), where similar tests can be used to evaluate emotional communication competence. Understanding, initiating and maintaining social relationships such as friendship (Baron-Cohen & Wheelwright, 2003), can be used as a guideline to prepare evaluations for the emotion regulation component. Understanding appraisals (Baron-Cohen, Leslie, & Frith, 1986; Lawson, Baron-Cohen, & Wheelwright, 2004) and intuitive physics (Baron-Cohen, Wheelwright, Spong, et al., 2001) can also be used to determine whether or not the system is capable of understanding cause and effect relationships, which is based on higher-level cognitive mechanisms.

Given the component model of empathy, it is beneficial to evaluate the model in both the component level and the system level. Component-level evaluation can provide an incremental assessment of the hierarchical nature of empathic behaviors. A poorly performing appraisal mechanism can be due to the malfunction of perception mechanisms or any other mediator, as well as a system-level misrepresentation of the components. Thus, evaluating the features separately before the system-level evaluation would help to provide useful insights into the nature of a problem. Nevertheless, this does not invalidate the necessity of providing a system-wide evaluation.

Moreover, studies have shown (Riek, Rabinowitch, Chakrabarti, & Robinson, 2009) that "human-likeness" and "believability" of the agents have a dramatic effect on feeling empathy towards the agent, especially when the situation evokes negative emotions (Rosenthal-Von Der Pütten et al., 2014). Considering the influential research of Reeves and Nass (1996) that demonstrated humans treat artificial agents as social actors, it is safe to assume that aesthetic decisions as well as perceived social traits of the agent would impact the perception of empathy during an interaction. Similarly, data-driven approaches might collect contradicting behavioral data from an empathic participant with regards to their interaction partner and the context. These social and aesthetic variables should be considered in the evaluation of empathic behavior.

2.7 Future Directions

Computational empathy research is still in its infancy. The challenges mentioned earlier need to be overcome in order to achieve competent empathic agents that can interact with humans in real time. Agreeing on better evaluation methods that are most suitable for measuring computational empathy is one of the most crucial issues that need to be solved. Novel evaluation metrics and questionnaires that are validated specifically for empathic agent research should be one of the central topics that the computational empathy community focuses on.

Another critical aspect for the future of computational empathy is to make use of the state-of-the-art research in affective computing and user modeling research. Recent advances in interactive systems, as well as the best practices of evaluation in both fields, can be translated into the implementation of the theoretical empathy models. Moreover, although human-level empathy does not often include the input from modalities such as skin conductivity, breathing, heartbeat, and the electrical activity from the brain, these modalities have been shown to be useful in gathering affective information. Greater use of these recent innovations in computational empathy research would surely contribute to the progress in the field.

Finally, the field of computational empathy is growing very rapidly and showing great promise for the future of AI. Our daily interactions are increasingly including artificial agents, which are evolving from mere tools to interaction partners and even companions. This rapid transition prompts important questions regarding the necessity of creating moral agents. Although the relationship between empathy and morality is a controversial topic (Decety & Cowell, 2014; Prinz, 2011), examining the effect of empathic capabilities on creating moral agents may provide significant insights.

2.8 Conclusion

Computational modeling of emotions has been useful for understanding, testing, and developing the theoretical framework of emotions in affective computing research. Computational modeling of empathy is a further development to this field. Research from philosophy, psychology, ethology and neuroscience provides a broad theoretical and empirical foundation for empathy modeling. A computational model of empathy should be grounded in this background research so that it can incorporate the full range of empathic behaviors. This paper is aimed to provide a holistic approach to empathy modeling that can account for the diversity of behaviors observed in various disciplines. We believe that a successful model and implementation of the empathic capacity not only would enhance our interaction with technological systems but also with each other as a society.

Chapter 3

A Computational Model of Empathy for Interactive Agents

This paper is published in Biologically Inspired Cognitive Architectures Journal: Yalçın, Ö. N. & DiPaola, S. (2018). **A computational model of empathy for interactive agents.** *Biologically Inspired Cognitive Architectures*, 26, 20-25.

Contributions: I was the main contributor to this paper. I was responsible for the conceptualization and development of the model as well as writing the original draft. Prof. DiPaola was supervising and reviewing the paper.

In this paper, I proposed a theoretical model of empathy for interactive agents following the theoretical foundations that I presented in the previous chapter. My model is mainly influenced by the evolutionary model of empathy by Preston and de Waal (Preston & de Waal, 2002). I used this influential model as a basis for the further development of my model of empathy to achieve the levels of empathic behavior in interactive artificial agents. I propose three hierarchical levels of components to achieve the spectrum of empathic behaviors as communication competence, emotion regulation and cognitive mechanisms. I further provided possible approaches to implement this model computationally by deriving state of the art methods from affective computing and social computing research.

3.1 Abstract

Empathy has been defined in the scientific literature as the capacity to relate another's emotional state and assigned to a broad spectrum of cognitive and behavioral abilities. Advances in neuroscience, psychology and ethology made it possible to refine the defined functions of empathy to reach a working definition and a model of empathy. Recently, cognitive science and artificial intelligence communities made attempts to model empathy in artificial agents, which can provide means to test these models and hypotheses. A computational model of empathy not only would help to advance the technological artifacts to be more socially compatible, but also understand the empathy mechanisms, test theories,

and address the ethics and morality problems the Artificial Intelligence (AI) community is facing today. In this paper, we will review the empathy research from various fields, gather the requirements for empathic capacity and construct a model of empathy that is suitable for interactive conversational agents.

3.2 Introduction

It has been argued that emotions have social and cognitive functions that is mandatory for an intelligent system (Minsky, 1991; Damasio, 1994). Empathy, as the capacity to relate another's emotional state (de Waal & Preston, 2017), is found to evoke altruistic and prosocial behavior, have a positive effect on the length of relationship, provide competence in communication and have a negative effect on aggressive behavior (Omdahl, 1995). These findings suggest that the study of empathy in artificial agents and especially in interactive systems can benefit from the development of a computational model of empathy in being able to evoke empathic emotions in the interaction partner and being able to act empathically towards the interaction partner Table 1. Studies conducted so far demonstrated that agents with the ability of showing empathy, a complex socio-emotional behavior, lead to more trust (Leite et al., 2014; Brave et al., 2005), have positive effect on the length of interaction (Bickmore & Picard, 2005), help reduce stress and frustration (Prendinger & Ishizuka, 2005; Burleson & Picard, 2004) and increase engagement (Coplan, 2011). Such a capability for computational systems would enhance the social interaction in educational applications, training environments, artificial companions, medical assistants and gaming applications, where initiating and maintaining a social interaction is of great importance.

As studies from neuroscience, ethology and psychology suggest, empathy is the result of complex interaction of lower and higher level cognitive processes (de Waal & Preston, 2017). Empathic capacity can be seen in a broad spectrum of behavior, which can be categorized as behaviors related to communication capacity, emotion regulation and cognitive mechanisms. A computational model of empathy must reflect the theoretical background and empirical findings. Existing models of empathy in artificial agents often fail to present a complete picture of empathic behavior or prefer to approach the problem by modeling only some components of empathic capacity, due to the complexity of the problem. However, the research outcomes so far show that it is worth studying empathy further to advance our understanding of empathy mechanisms, and to provide better interaction techniques into human-computer interaction research. In this paper, we propose a model of empathy which is suitable for conversational agents and provide an overview of techniques that can be used to implement such an agent.

3.3 Theoretical background on empathy

Since its philosophical foundations, empathy has been a research topic in several research fields such as clinical psychology, social psychology, aesthetics, ethics and neuroscience (see Coplan and Goldie (2011) for an extensive overview of empathy research). The behaviors assigned to empathy gathered in these various fields are mirroring, affective matching, sympathetic concern, consolidation, theory of mind, perspective taking and projection of self (de Waal & Preston, 2017; Coplan, 2011; Batson, 2009; Leiberg & Anders, 2006). These behaviors are usually categorized as affective or low-level empathy and cognitive or high-level empathy.

Table 3.1: Summary of empathy components, the mechanisms involved and the empathic behavior they are linked with.

Empathy Component	**Mechanisms**	**Empathic Behavior**
Communication Competence	Emotion Recognition Emotion Expression Emotion Representation	Mirroring Affective Matching
Emotion Regulation	Features of the observer Features of the relationship	Empathic Concern Consolation
Cognitive Mechanisms	Emotional Appraisal Theory of Mind Simulation Theory	Alturistic Helping Perspective-Taking

Affective empathy is defined as the automatic and often unconscious mimicking of other's emotional response and includes mirroring and affective matching behaviors. Cognitive and high-level empathy is considered as the ability to understand the other's emotional and mental states by using cognitive mechanisms such as perspective taking and theory of mind (ToM). There is disagreement on how these behaviors are related, and whether empathy is a discrete or continuous phenomenon (Coplan, 2011). However, research efforts suggest the levels of empathy are interconnected (Preston & de Waal, 2002; Coplan, 2011). De Waal and Preston (2017) suggests the components of empathy are built on top of each other due to evolutionary mechanisms (see Figure 1). According to this view, perceiving the behavior of another in one's own representation is automatic in and based on the perception–action mechanism (PAM). This mechanism is the essential for affective and cognitive empathic behavior.

Different theoretical approaches to model empathy agree on a set of cognitive and behavioral capabilities required for empathic behavior, which we will call components. Understanding these main components and the underlying processes is crucial for simulating empathy in AI based interactive agents, regardless of the chosen theoretical model. These components can be categorized as communication competence, emotion regulation and cognitive mechanisms. Similar components have been suggested by Paiva and colleagues (2017),

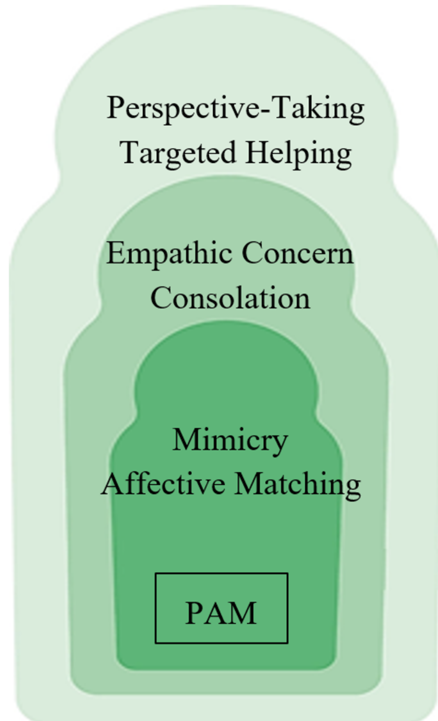

Figure 3.1: Russian Doll Model of Empathy from De Waal and Preston (2017).

and Boukricha and colleagues (2013) as empathic responses, empathy modulation and empathy mechanism. Their categorization of empathic capacities does not include the perception of emotions as an essential communication capability, but rather as one of the mechanisms of empathy.

Another categorization by Davis's (1994) refers to antecedents, processes and empathic outcomes that resemble this approach. Antecedents, refer to the biological capacities and individual differences that affect the strength of empathic response, similar to emotion regulation and modulation parameters. Processes refer to the type of cognitive or noncognitive processes that result in the empathic response, such as mimicry, association and role taking. Empathic outcomes can be intrapersonal and interpersonal outcomes that shows the empathic behavior presented. According to this view, intrapersonal outcomes include parallel and reactive outcomes as well as non-affective judgements. Interpersonal outcomes are helping behavior, aggression or social behavior.

3.4 A proposed model of empathy

Following the theoretical foundations and research on empathy from various fields, we can conclude that the empathic capacity should contain a broad spectrum of behaviors that we can categorize as communicative ability or emotion, regulation of emotions and the cognitive mechanisms related on the appraisal of the emotional situation. A model of empathy for interactive agents should include these capabilities that are linked to each other to express different levels of empathic behavior. Our proposed model is therefore composed of three-level structure, as shown in Fig. 2, is aimed to capture the link between empathic behavior and biological capabilities that are needed to satisfy them.

The hierarchical structure of the components in this model resonates with the layered representation of the Russian doll model of empathy (de Waal, 2010), where each layer is built on top of the mechanisms of the previous layer. In the following sections we will provide a description of each component.

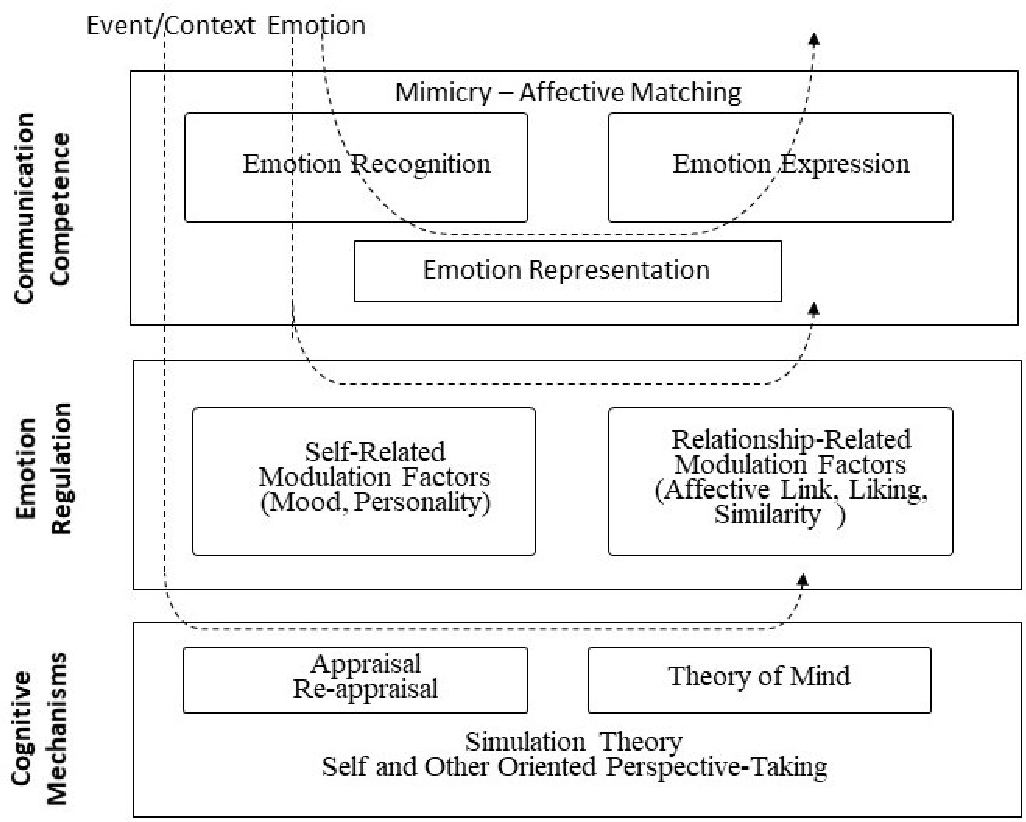

Figure 3.2: The proposed model of empathy with hierarchical components.

3.4.1 Communication competence

The definitions and models of empathy suggest an underlying mechanism required for all levels and intensities of empathic behavior, where without the successful recognition and expression of an emotion it is impossible to show any type of empathic response towards it. The concept of emotional intelligence (Salovey & Mayer, 1990) in psychology is addressing this distinct capacity of perceiving, accessing and generating emotions to be used in assisting thought, reasoning and understanding. This includes identifying and expressing emotions in other people as well as art pieces. Similarly, Scherer (Scherer, 2010a) focuses on Emotional

Production and Recognition Competence while referring to distinct skills that are underlying factors for any type of emotional behavior. Preston and de Waal (2002) points out that neurological disorders such as autism, sociopathy and prefrontal damage suggests that the lack of empathy is linked to the disfunction of recognition and expression of emotions. De Vignemont & Singer (2006) mentions these the communication capacities and emotional repertoire of the empathizer effects the intensity of the empathic response. Therefore, it is mandatory for any computational approach of empathy to include the emotion recognition and emotion expression capabilities, which we will call communication competence. The lowest level of the hierarchy of the empathy model is the affective communication competence, which includes emotion recognition, emotion expression and emotion representation capabilities.

Low-level or affective empathic behavior, which includes motor mimicry and affective matching arises from the affective communication competence. Mirroring behavior such as facial and motor expression, yawn contagion and emotional contagion are the low-level empathic behavior are the products of a basic perception-action mechanism that allows humans to resonate the observed actions with their own representations and actions. According to the Russian doll model of empathy by de Waal (de Waal, 2007), the motor mimicry and emotional contagion are at the foundation of any empathic behavior and are based on perception-action mechanism (PAM). PAM suggest humans can understand emotions due to the mapping of other's emotional state onto their own representations of those emotional states. Therefore, the affective communication competence requires the recognition of emotions, expression of emotions and the internal representation of those emotions within the agent.

3.4.2 Emotion regulation

Another main component of empathy is the regulation of the empathic emotions that is related to a range of psychological, social and cultural factors. De Waal and Preston (2017) mention that perception-action mechanisms (PAM) is not limited to perform in the low-level mimicry and emotional contagion stages that only requires automatic recognition and expression mechanisms. They argue that more complex forms of empathy require the interaction of PAM with regulatory processes such as attention, experience, as well as similarity, relatedness, familiarity and social closeness (de Waal & Preston, 2017). Scherer (2010a) defines Emotion Regulation Competence as the capacity to modify the "raw" emotion by a set of automatic and conscious monitoring mechanisms.

Davis (1994) mentions antecedents as biological and individual differences that result in modulating the strength of empathic response. This model was used by de Vignemont and Singer (2006), who illustrate these modulatory functions as features of the emotions (valence, intensity, saliency, primary vs. secondary emotions), social link between target and empathizer, context, and the characteristics of the empathizer (mood, personality, gender, age). Hoffman (2000) further adds familiarity bias that includes in-group, friendship and similarity biases. Coplan (2011), states the inhibition of these mechanisms is required for other-oriented perspective taking.

Paiva and colleagues (2017) provide a categorization of these factors (referred as empathy modulation factors) in empathy research: features of the observed emotion, relationship between the observer and the target, situation/context and features of the observer. They mention the shared concepts between the emotion research and empathy research regarding these regulatory factors and state how the OCC model of emotions introduce a set of variables that influence the intensity of the emotions.

The regulation mechanisms include the self-related and other related emotion regulation factors. Self-related factors include characteristics of the empathizer that are independent on the observed emotion or situation such as mood and personality. These factors alter the expressed emotion in different timescales as well as effecting the representation of emotions, the attention mechanisms towards the observation and can have an impact on how the observed situation will be interpreted in the cognitive mechanisms. Relationship-related regulation factors are factors that are dependent on the link between the observer and the target. Liking, similarity, social link, and familiarity biases are some of these factors that effect the expression of the emotional behavior. Context is another important parameter that may affect the expressed empathic emotion.

3.4.3 Cognitive mechanisms

High-level or cognitive empathic behavior requires a set of cognitive capabilities such as appraisal and re-appraisal of the situation, theory of mind and mental simulation. These capabilities form the top level in the hierarchy of the empathy model and depends on the lower levels in the model. Cognitive mechanisms can be affected by the emotion regulation factors and mirroring mechanisms. Therefore, higher level empathy mechanisms such as perspective taking, and targeted helping require the inhibition of the effects of the lower level components.

In cognitive theories of emotion, appraisals are the evaluations of events happening in the environment of the agent based on the impact of that situation (Lazarus, 1991a). Appraisals can be made in accordance to user's cognitive functions such as intentions, plans and goals as well as emotions and social values. The cognitive approaches to emotions (Arnold, 1960) state that, for an event to invoke emotional response, it should affect a person based on the goals of a person. The OCC (Ortony et al., 1990) theory of emotion belongs to this category of emotional models, which is frequently used in virtual agent research. The agency component of this theory is very suitable for simulating human cognitive processes, which is a key for higher level/cognitive empathy. However, the appraisal mechanisms mentioned by the emotion research that ties the goals and values of the organism to cognitive function cannot be enough for high-level empathy where perspective taking is required (Coplan, 2011). Higher cognitive processes such as theory of mind, perspective taking, and projection of self is required for higher level empathy functions (Goldman, 2012).

Omdahl (1995) demonstrates how cognitive appraisal of an affective situation allows people to arrive at assessments about the situation not only on self-directed perspective taking mechanisms but also on other-directed perspective taking mechanisms by re-appraisal of the situation. Following the work on appraisal theory (Scherer, 1982) and semantic network-theory (Bower, 1981), Omdahl (1995) demonstrated that reasoning about other's emotions and perspective taking are related to appraisal and re-appraisal of the situation. This notion of achieving perspective taking by reappraisal is also proposed in de Vignemont & Singer (2002).

3.5 Empathy in conversational agents

In her extensive study on the link between cognitive appraisal theories and empathy, Omdahl (Omdahl, 1995) argues about the special need to examine verbal information in empathy research as it is the primary source of understanding the context of affective information. This idea suggests that conversational agents can be beneficial in empathy research, as they can be used to provide additional context to affective and social signals. However, as empathy refers to a range of phenomena with complex interactions, the empathy models and definitions used in conversational agent research do not always match the theoretical background.

Conversational interaction has a central role in human-to-human communication, where information from multiple sensory channels are received and sent simultaneously. Embodied conversational agents (ECAs) are agents that can interact with users with a multimodal, situated (and often anthropomorphic), and real-time interaction to emulate a similar experience of human-to-human conversational interaction (Cassell, Bickmore, Campbell, & Vilhjálmsson, 2000). There is a tendency in ECA research to refer to sympathetic (D'mello & Graesser, 2012; Looije, Neerincx, & Cnossen, 2010), emotionally supportive (Brave et al., 2005) or compassion behavior (Looije et al., 2010) as empathy, which only reflects narrow fragments of empathic behavior. Others focus on the low-level empathic functions such as mirroring (M. Smith, 2011) and affective matching (Lisetti et al., 2013) in conversational agents, that may provide a basis for higher-level phenomena. Some researchers attempt to model higher-level empathic behavior, such as reactive empathy in Davis (1994), however do not provide any emotion-regulation or appraisal mechanisms to justify agent actions (Moridis & Economides, 2012). Even though these approaches do not reflect the full array of empathic reaction, they provide a good foundation for building up a more complex empathy model.

More complete models of empathy have also been studied in ECA research. The EMMA framework (Boukricha et al., 2013) uses a continuous representation of empathy that includes modulation mechanisms that uses mood, liking and familiarity factors to adjust the expressed empathic emotion, where empathy is defined as affective matching. Their theoretical model includes higher level empathy mechanisms however, was not implemented in their agent framework. Another example is the CARE framework (McQuiggan et al., 2008), the agent learns empathy by observing the interactions between humans in a treasure hunt game called Crystal Island. Their system can track intentions, age, gender, affective state and goal of the users. The system learns to show parallel or reactive empathy depending on the context and attributes of the target.

Furthermore, some research efforts are tackling similar problems and generate matching models of emotion without mentioning empathy. GRETA's (Clavel & Callejas, 2016) motor resonance module can be regarded as a low-level empathy component that allows for mimicry where the agent-mind takes the goals, beliefs and social relationships into account to have a long-term decision, which can be regarded as a higher-level regulation mechanism. A more complex form of empathic behavior can be found in the user adaptation applications in dialogue systems (Jokinen, 2003), where the aim is to model the user intentions and state to predict and adapt to the user behavior (Jokinen, 2010). The TRAINS-92 system (D. R. Traum, 1993) was aimed to predict user intentions to adapt the dialogue state accordingly, which can be seen as a foundation for high-level empathy.

The general challenges in implementing empathy models in ECA research are reflecting the multi-level nature of empathic behavior, integrating the context, and automatization of the behaviors (Paiva et al., 2017). However, the growing number of research efforts in empathic agent research and technological advances suggest these challenges might be overcome in near future. The fields of affective computing and social computing as well as dialogue management research provide valuable insights on how to implement empathy based on emotional communication, regulation and cognitive capacities.

3.5.1 Affective computing

It was argued that emotional intelligence, a set of social and emotional skills to recognize, regulate and express emotions, is crucial for a successful communication, coping with the social life and empathy (Bar-On, Tranel, Denburg, & Bechara, 2004). The ability to detect, understand and exhibit affective signals is therefore said to be indispensable for successful interaction (Goleman, 2005). These capabilities lay the foundation for any level of empathic behavior and is crucial for understanding others. The research area for building systems that can recognize and express emotions is called affective computing (Picard, 1997).

Affective computing research helped to develop effective techniques for emotion recognition and expression in artificial conversational agents, as well as testing and creating models of emotion (Gratch & Marsella, 2004) and simulating computational agents for human-computer interaction (HCI) applications (Pantic, Sebe, Cohn, & Huang, 2005; Jaimes & Sebe, 2005). The body of work has proved the importance of adapting to the user's moods, preferences, attention and intentions. Interaction systems that can recognize and react to the affective states of the users are more likely to be perceived as natural, trustworthy and efficient (Pantic et al., 2005).

Affect communication with ECAs is crucial to determine user's emotions and attitudes to have an adaptive interaction; and being able to express them properly to have more meaningful and natural interaction (J. Bates et al., 1994). It was found that detection of sentimental cues during dialogue is valuable to achieve long-term interaction between humans and ECAs (Pecune, 2013). Emotions can also be viewed as a control mechanism for the range of actions that the agent can choose from (Wilks, Catizone, Worgan, & Turunen, 2011).

Researchers use two main approaches to represent emotions: categorical emotion recognition (such as fear, anger, happiness) (Ekman, 1992) or dimensional approach (valence and arousal space) (Osgood, May, Miron, & Miron, 1975). Valence represents the pleasantness of a stimuli between positive and negative scale, where arousal shows the intensity of the emotion. Another approach on appraisal theories suggests that emotions are triggered due to the assessment of the events based on the goals and abilities of the agent (Scherer, Bänziger, & Roesch, 2010). The OCC model of emotions (Ortony et al., 1990), Scherer's appraisal model (Scherer, 2001) or EMA (S. C. Marsella & Gratch, 2009) are the most used appraisal models in ECA research.

Clavel and colleagues mention (2013) the choice of computational and theoretical models should be related to the capabilities and goals of the embodied conversational agent. Dimensional analysis has been widely used in detecting negative-positive emotions and model agent behavior (C. Smith et al., 2011). Discrete approaches are used to detect certain emotional categories such as user frustration or attention (D'mello & Graesser, 2012) that can be beneficial to track application-specific emotions. Appraisal based models can be used to guide the dialogue, model agent behavior (El-Nasr, Yen, & Ioerger, 2000) and to model the user's affective state (Jaques & Vicari, 2007).

These techniques used in emotion recognition, representation and expression can be easily used as a foundation for an empathic conversational agent and can be used to generate low-level empathic behavior and inform higher-level components of empathy. Moreover, cognitive models of emotion such as appraisal models can be used to reason about the experienced situation, react to the situation based on its appraisal and the reappraisal of the situation are fundamental for cognitive mechanisms of empathy.

3.5.2 Social computing

A key feature of the social communication is the detection of social signals that helps us to predict the behavior of the interactive partner and act accordingly. Social signals are defined as "a communicative or informative signal that, either directly or indirectly, provide information about social facts, namely social interactions, social emotions, social attitudes, or social relations" (Vinciarelli et al., 2012, p. 4). A successful social interaction requires the detection of the correct non-verbal social signals from machine sensors, reasoning about these signals in a dynamic way and expressing social signals based on that reasoning (Mairesse & Walker, 2010). The empathic behavior can be regulated based on self-related and relationship-related social signals. In social computing research, the personality of an agent can influence the expression of behaviors and can be used to capture the personality of the target to compute the similarity between the observer and the target.

It is possible to capture the engagement, stress and interest levels of the users, which is useful for understanding user preferences and providing feedback to the system about its performance (Pentlan, 2005). Moreover, recognition of the patterns in social signals can be used for modeling personality traits. The most adopted personality model in social computing research is the Big 5 personality model (P. T. Costa & MacCrae, 1992) that represents personality in openness, consciousness, extraversion, agreeableness and neuroticism dimensions. However, it is a challenge to integrate the effect of context, beliefs, goals and intentions of the agent with these social signals (Wang, Carley, Zeng, & Mao, 2007). Automated methods have been proposed for personality recognition from conversational data. A recent approach on text-based conversational models uses Speaker-Addressee model using LSTM network approach to generate consistent responses that matches personalities which is represented as a vector embedding (Li et al., 2016). This vector representation holds the information about dialect, age, gender and other personal information that will influence the content and style.

Moreover, as a feature of cognitive mechanisms, simulation of the target's beliefs, desires and intentions within the observer is required for perspective-taking. User modeling techniques used in social computing research can provide a basis for internalizing the other's mental states within the agent. Kobsa and Wahlster (1989) define the user model as a knowledge base that contains assumptions about the user that might be relevant to the behavior of the agent, which are separate from the rest of the agent's knowledge. Such a user model dynamically that consists of the general knowledge, beliefs and the goals of the user can allow for higher-level empathic reasoning and perspective taking. An adaptive capability for empathic conversational agent should address the issues of how to recognize user traits, deciding on which of those traits to be reasoned with and expressing consistent responses based on that decision.

3.5.3 Context in conversational systems

Context can provide valuable information about the cause and effect relationship between the event and the emotional effect it triggers in the target. Verbal channel can offer important cues for decoding of the situation and empathy. Linguistic context has been found to be vital in dialogue management (Pickering & Garrod, 2004), recognizing emotions (Calvo & D'Mello, 2010; Picard, 2014; Cambria, 2016), understanding social behavior (Vinciarelli et al., 2012) and for multi-modal interaction (Pantic et al., 2005; Jaimes & Sebe, 2005). Same emotional gestures might signify different meanings based on their context (Ekman, 2004). Understanding and representing context of the verbal information is therefore crucial for an empathic ECA.

In dialogue systems, the context can be extracted by using methods as simple as a keyword detection mechanism or more advanced techniques as natural language modeling and vector representation. The recent advances in deep learning techniques made it possible to create a distributed representation of a given text (Bengio, Ducharme, Vincent, & Jauvin, 2003; Mikolov, Chen, Corrado, & Dean, 2013) which can be used to understand the context of an utterance.

The traditional conversational systems follow a modular approach (D. Traum, 2017; Jokinen, 2010) which consists of Natural Language Understanding (NLU) module that processes the input from the user, the Dialogue Manager (DM) that updates the internal state to select a proper response given that input and Natural Language Generation (NLG) module that generates and outputs a proper utterance according to that internal state. Recently, advances in Deep neural networks (DNNs) has been very successful in many applications of NLP, speech recognition and response generation. However, these approaches still need to be integrated into ECA frameworks.

3.6 Conclusion

Empathic capacity, as the capacity to relate and react to another's emotional state, consists of emotional communication competence, emotion regulation and cognitive mechanisms that result in a broad spectrum of behavior. A computational model of empathy should embody these capabilities to present a complete array of empathic behavior. Existing models and implementations of empathy in interactive agents including conversational agents often portray an incomplete picture of empathy or represent a binary classification of empathy. We propose a hierarchical approach to model empathy in interactive agents, that can be implemented by using some of the existing techniques in affective computing, social computing and dialogue research. Although each of these research areas provide advanced techniques that cover individual factors and components of empathic capacity, successful integration of them to reflect the hierarchical multi-leveled structure of empathy requires further investigation.

Chapter 4

Evaluating Empathy in Artificial Agents

This paper is accepted in the Affective Computing and Intelligent Interaction Conference 2019: Yalçin, Ö. N. (2019). **Evaluating Empathy in Artificial Agents**. In *8th International Conference ACII*, (pp. xx-xx).

Contributions: I was the main and sole contributor to this paper.

In this paper, I propose a systematic approach to evaluate computational empathy in computational systems. The novel research area of computational empathy lacks specific tools to evaluate the levels of empathic behavior. Following the well-established methods of evaluating empathy in humans and evaluation metrics for interactive agents from HCI research, I provide recommendations for evaluation methods for interactive artificial agents. These recommendations are aimed to initiate the dialogue for the collective development of better metrics to evaluate empathy in artificial agents. In the following chapters, I will apply the proposed approach to the evaluation of the empathic embodied conversational agent that will be implemented.

4.1 Abstract

The novel research area of computational empathy is in its infancy and moving towards developing methods and standards. One major problem is the lack of agreement on the evaluation of empathy in artificial interactive systems. Even though the existence of well-established methods from psychology, psychiatry and neuroscience, the translation between these methods and computational empathy is not straightforward. It requires a collective effort to develop metrics that are more suitable for interactive artificial agents. This paper is aimed as an attempt to initiate the dialogue on this important problem. We examine the evaluation methods for empathy in humans and provide suggestions for the development of better metrics to evaluate empathy in artificial agents. We acknowledge the difficulty of

arriving at a single solution in a vast variety of interactive systems and propose a set of systematic approaches that can be used with a variety of applications and systems.

4.2 Introduction

Emerging technologies continue to change the ways in which we interact with computers. Computational systems are evolving from being mere tools to assistants, trainers and companion agents. All of these new roles assigned to these systems highlight the importance of embodying these agents with social and emotional capabilities. The advances in computational interaction techniques allowed for the development of emotionally sensitive, perceptive, socially situated and expressive agents. One of the novel and exciting addition to these behaviors is empathy, as a complex socio-emotional behavior.

Empathy can be defined as the capacity to perceive, understand and respond to others' emotions in a manner that is more suitable to those perceived emotions than one's own (Preston & de Waal, 2002). The long history of empathy research with the contribution of many disciplines (e.g. philosophy, psychology, neuroscience, ethology) resulted in a diverse set of definitions and behaviors assigned to empathy (see (Coplan & Goldie, 2011) for the history of the field). Behaviors such as mimicry, affective matching, consolidation and perspective-taking are assigned to empathic ability in humans (Coplan, 2011), which are crucial to initiating and maintaining social relationships. Following the centuries-old research that shows the importance of empathy in social interactions, computational empathy emerged as a novel field to equip artificial agents to show empathic behavior during their interactions.

With the recent developments in the capabilities of computational systems, the computational modeling of empathy has gained increasing attention in the last decade. Research on computational modeling of empathy have shown that empathic capacity in interactive agents lead to more trust (Brave et al., 2005; Leite et al., 2014), increase the length of interaction (Bickmore & Picard, 2005), help coping with stress and frustration (Prendinger, Mori, & Ishizuka, 2005) and increase engagement (Leite et al., 2014) (see (Paiva et al., 2017) for a review of the field). These findings suggest that agents with empathy could enhance the social interaction in educational applications, artificial companions, medical assistants and gaming applications. Equipping artificial social agents with empathic capabilities is, therefore, a crucial and yet challenging problem.

Part of this challenge arises from the lack of evaluation methods to measure empathic behavior in artificial agents, which would allow for a systematic assessment of the steps need to be taken to model the components of this complex socio-emotional phenomenon. Although various fields provide well-established evaluation methods to measure empathy in humans, it is not clear how to translate these methods to evaluate computational systems. Empathy research in psychology provides validated methods to measure empathy levels in a person (Davis, 1983; Baron-Cohen & Wheelwright, 2004; Lawrence et al., 2004). These methods often require a first-person report of certain behavioral traits based on subjective questionnaires. This poses a challenge for artificial empathy work in virtual agents, as subjective measurements cannot be used in artificial entities (Paiva et al., 2017). Moreover, behavioral tests from psychology and neuroscience often rely on physiological signals such as neural activity, heart rate and skin conductance, which cannot be used for machines. Other methods may require the observation of experts in the field, which is hard to automate.

Furthermore, agents that have varying levels of interactive capabilities and application goals can have different effects on the perception of the agent and interaction. The characteristics of the agent (e.g. aesthetics, embodiment) as well as non-functional properties (e.g. fluency, response time) can affect the evaluation of empathy, as much as the empathic functionality. A conversational agent in a text-only environment such as chatbots would require emotion perception and expression using different modalities than an embodied conversational agent. In that sense, every additional capability of the agent would contribute to the evaluation of the interaction and the system. Similarly, the application areas may enforce varying levels of expectations in terms of empathic behavior. A medical assistant may require more sympathy, where a personal trainer would focus on pushing boundaries of the interaction partner. This diversity highlights the importance of a set of evaluation methods that allow for flexibility, instead of focusing on reaching to a single solution.

This paper aims to provide recommendations on how to systematically evaluate empathy in artificial agents in a variety of contexts and capabilities. We propose to approach this problem by focusing on best practices in both the empathy research in humans and the human-computer interaction (HCI) research. To achieve this, we will examine the methods of evaluation in empathy of humans to draw conclusions on how to adapt the key concepts to computational empathy. We propose to use system-level and feature-level evaluations to systematically list the factors that contribute to the evaluation of empathy. By providing a checklist of these factors, we aim to initiate the discussion towards creating a common ground on the evaluation of empathy in artificial agents.

In the following sections, we will explore the evaluation methods on empathy in humans (Section 4.3) and try to apply this know-how into evaluating empathy in artificial agents using system-level and feature-level evaluations (Section 4.4). We will conclude with a discussion of challenges and a call for collective action to work towards new evaluation methods in this new and exciting field.

4.3 Evaluation of Empathy in Humans

Empathy research from many disciplines have developed definitions and models for empathy that resulted in a variety of capabilities assigned to empathic behavior (Coplan, 2011). Capabilities such as mimicry, affective matching (emotional contagion), sympathy (empathic concern), altruistic helping (consolidation) and perspective taking are assigned to empathic behavior by scholars (Coplan & Goldie, 2011; de Waal & Preston, 2017; Batson, 2009). How many of these behaviors should constitute empathic behavior and how they are connected are still highly debated topics in empathy research. Following the differences in the definitions and models of empathy, the evaluation metrics developed to measure empathic behavior may vary dramatically.

Some definitions of empathy separate a number of these capabilities as affective and cognitive empathy (Omdahl, 2014). According to this view, affective empathy refers to the relatively automatic emotional responses to other's emotions. Behaviors related to affective empathy can be listed as mimicry, affective matching and empathic concern (Eisenberg & Strayer, 1987). On the other hand, cognitive empathy includes behaviors that require the understanding of another's emotional state and behaviors. Behaviors such as consolidation, perspective-taking and altruistic helping are said to originate from the involvement of cognitive mechanisms during the processing of the other's emotional situation (de Waal & Preston, 2017).

Contrary this dual view of empathy that separates affective and cognitive processes, a unifying view of empathy is gaining attention as an alternative. These recent models and definitions of empathy suggest a more multi-dimensional approach where both affective and cognitive empathy are interconnected with a variety of processes that results in individual differences (Davis et al., 1980; Preston & de Waal, 2002; Hoffman, 2001). One of the most prominent views of empathy called the Russian Doll Model of Empathy (de Waal, 2007) suggests hierarchical levels of affective and cognitive capabilities are connected through evolutionary mechanisms. According to this model, processes such as emotional communication capabilities (recognition and expression), emotion regulation, appraisal processes and theory of mind are considered to be the foundational mechanisms that allow the levels of affective and cognitive empathic behavior (de Waal & Preston, 2017). It is also suggested that the individual differences between the empathic responses of people are related to the factors that affect the outcome of these processes (Davis et al., 1980). For instance, the recognition of emotions depends on the intensity of the perceived emotion, where the regulation of the emotion would depend on the familiarity between individuals as well as the features of the observer (mood, personality) (de Vignemont & Singer, 2006).

The evaluation metrics to measure empathic capacity in humans varies depending on the definition and the capabilities that are assigned to empathic behavior. Evaluation metrics that follow the categorical view of empathy usually assess the affective or the cognitive aspects of empathic behavior as separate constructs. On the other hand, evaluation metrics that follow the multi-dimensional view focuses on defining and evaluating the levels of processes and behaviors that determine the extent of empathic behavior. Although both of these approaches received criticism by some researchers that suggest empathy should only involve the higher level processes (Coplan, 2011), it is useful to focus on a broader view of empathy for to arrive at a comprehensive framework to evaluate empathy in artificial agents. Following this notion, we will explore the evaluation methods by focusing their adaptability to computational empathy research.

In this section, we will give an overview of the well-established measurements for empathic behavior, while categorizing them in terms of the focus of empathic behavior and the method of delivery. Evaluation metrics can target a variety of levels of empathic behavior with different levels of granularity and abstraction. Some of the evaluation methods are designed to measure empathy as a single comprehensive construct and aimed to derive a single value that would indicate the global empathy. Others focus on multiple features or a subset of those features that underlie empathic behavior, such as the affective and cognitive capabilities that are mentioned earlier. Therefore, we will categorize the evaluation metrics according to the level of granularity they are aimed to quantify: global empathy and components of empathy.

Moreover, the method of delivery for these evaluations can be categorized as self-report, observational and physiological approaches. Physiological approaches include measurements of brain activity or autonomic nervous system measures (heart rate, skin conductance, breathing rate). Methodologically, there is no direct way of applying the physiological approaches to virtual agents. Therefore, we will not cover them in our paper (see (Neumann, Chan, Boyle, Wang, & Rae Westbury, 2015) for a review of these approaches). Self-report measures usually include surveys/questionnaires that rely on the individual's assessment of their behavior. Observational methods can include the behavioral tests and perceived empathy measures. Behavioral methods rely on performance tests based on experimental stimuli. These methods are often used to assess the components of empathy in humans and aimed to indicate deficiencies. Lastly, perceived empathy metrics are questionnaires that require an observer's assessment of an individual's behaviors. These can include expert observations on the subject's behaviors as well as a second or third person account of a non-expert. In the following sub-sections, we will give detailed examples on the well-established evaluation metrics on each of these methods within global and component-based evaluation metrics.

4.3.1 Evaluating Global Empathy

Measures of global empathy are aimed at quantifying a single value that would indicate the strength of empathic capability as a broader concept. Many researchers have attempted to develop self-report measures of empathy based on the definitions and capabilities they assigned to the term. Most of these methods focus on the evaluation of empathy as a whole, while others focus on the specific factors that add up to the global empathic behavior.

One of the earliest self-measure of empathy is Hogan's Empathy Scale (ES) (Hogan, 1969) that is mostly used to assess cognitive empathy with 64 true-false statements taken from the standard psychological scales. This questionnaire was intended to examine the relation of empathy with moral and socially appropriate behavior. It was criticized by later works that it is better suited for the evaluation of social skills in a broader sense rather than a specification of empathy (Davis, 1983; Baron-Cohen & Wheelwright, 2004). Moreover, the low scores of the validity and reliability of the scale resulted in a continuous decrease in the use of this scale as a valid measure of empathy (Froman & Peloquin, 2001). However, this attempt encouraged researchers to investigate further and examine the development of a more suitable evaluation of empathy.

A frequently used example is Davis's Interpersonal Reactivity Index (IRI) (Davis, 1983), is a 28-item scale for multi-dimensional measurement of empathy with four sub-scales: perspective-taking, empathic-concern, fantasy and personal distress. However, there have been some discussions around the appropriateness of this scale to measure empathy. Firstly, it was argued that the questionnaire may capture behaviors broader than empathy (Baron-Cohen & Wheelwright, 2004; Spreng, McKinnon, Mar, & Levine, 2009), such as imagination (e.g. item 1 "I daydream and fantasize, with some regularity, about things that might happen to me") and emotional control (e.g. item 10 "I sometimes feel helpless when I am in the middle of a very emotional situation"). An adaptation of IRI to exclude the “fantasy" subscale was later adopted as Feeling and Thinking Scale (Garton & Gringart, 2005). It was also suggested that the “personal distress" sub-scale mostly measures anxiety towards distressing situations in general and does not relate to the core functions of empathy. Moreover, some researchers suggested the further refinement of this scale due to the correlation between these sub-scales (Spreng et al., 2009).

The Empathy Quotient (EQ) (Baron-Cohen & Wheelwright, 2004) is one of the most accepted self-report scales that is validated by numerous studies (Lawrence et al., 2004). Authors define empathy as “the drive to identify another person's emotions and thoughts and to respond to these with approapriate emotion" (p.361). This test is aimed as a clinical screening tool for adults with Autism Spectrum Disorders. In contrast with other self-report questionnaires of empathy, authors did not differentiate between affective and cognitive empathy. They aim to capture empathy in a broader sense where both levels have very interrelated capacities. This questionnaire includes 60 items with 40 empathy-related and 20 filler questions answered with a 4-point Likert scale that scores agreement with the statements. Example questions from the EQ are “I am good at predicting how someone will feel" and “Seeing people cry doesn't really upset me." The questionnaire scores are shown to correlate with autism and gender differences (Baron-Cohen & Wheelwright, 2004).

A recent attempt to further examine and combine these self-report measures uses factor-analysis to reach to a brief and reliable measurement of empathy is called The Toronto Empathy Questionnaire (Spreng et al., 2009) (TEQ). Authors gathered a total of 142 items from several self-report empathy questionnaires such as IRI, ES, BEES, QMEE, AQ as well as empathy questionnaires for specific populations such as Jefferson Scale of Physician Empathy (Hojat et al., 2001), Nursing Empathy Scale (Reynolds, 2017) and Japanese Adolescent Empathy Scale (Hashimoto & Shiomi, 2002). Authors used these items to refine a final set of 16 items that are found to be most correlated with Empathy scores compared to other questionnaires. Responses are made with 5-point Likert scale items that show agreeableness of the statements. This questionnaire is a shorter alternative to the EQ with high internal consistency, validity and reliability scores.

A similar approach is taken in the Questionnaire of Cognitive and Affective Empathy (QCAE) (Reniers, Corcoran, Drake, Shryane, & Völlm, 2011), which is derived from EQ, ES, IRI and the Impulsiveness-Venturesomeness-Empathy Inventory (Eysenck & Eysenck, 1978). Authors finalized a 31-item questionnaire that measures cognitive and affective empathy, as the name suggests. The QCAE consists of five sub-scales, where two sub-scales are related to cognitive empathy (perspective-taking, online sumulation), and three of them are related to affective empathy (emotion contagion, proximal responsivity and peripheral responsivity).

Other methods focus on the evaluation of the perception of empathic behavior. These measurements provide a second and third person perspective on an individual's empathy with questionnaires. Jefferson Scale of Physician Empathy is developed to evaluate empathy as a predominantly cognitive attribute (Hojat et al., 2001). These components are "communication", "understanding" and "cognition", are focused on the cognitive empathy, rather than the affective empathy that was mentioned earlier (see Section 4.3). The scale consists of 20 items with a 7-point Likert type scale on agreement with the statements. Another example for the perceived empathy methods is the Consultation and Relational Empathy (CARE) Questionnaire (Mercer, Maxwell, Heaney, & Watt, 2004). CARE was developed to measure "relational empathy" that focuses on the social function of empathy. It consists of 10 statements that start with "How was the doctor at ..." and scored by the patient by using a 5-point likert scale from "poor" to "excellent". The items include "Fully understanding your concerns", "Showing care and compassion" and "Being positive". Some of these items show significant overlap with self-report questionnaires such as the IRI.

4.3.2 Evaluating Components of Empathy

An alternative approach to the evaluation of global empathy is the evaluation of specific components that are required for empathic behavior. These evaluations are usually done by testing behavioral and cognitive abilities to detect the deficits and abnormalities in the behavior. Components such as emotional communication (recognition and expression), emotion regulation, appraisal and perspective taking are usually targeted in these evaluations as critical mechanisms for levels of empathy.

The "reading the mind in the eyes" test (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997) was one of the first examples of these behavioral tests. This test was aimed to be used as a screening test for adults or children with Aspergers Syndrome, who are considered to have a deficit in empathy. The revised version of this test consists of 36 photographs of the eye-region of the face that shows different emotional expressions (Baron-Cohen, Wheelwright, Hill, et al., 2001). The participants are presented with these photographs and are asked the most appropriate word to describe “what the person in the photograph is thinking or feeling" (p.241) among four words that are presented. The target terms include words that show mental states such as “thoughtful", “interested" or “fantasizing", as well as words that relate to the emotional state such as “upset", “nervous" or “hostile". The test results indicate the ability to perceive social and emotional cues where a lower score is associated with a broader set of phenomena than just measuring empathy.

Similarly “reading the mind in the voice" test (Golan et al., 2007) and “reading the mind in films" test (Golan, Baron-Cohen, Hill, & Golan, 2006) are aimed to measure the ability to detect socio-emoitonal cues in voice and movie stimuli respectively. The “voice" test uses segments of dialogue taken from dramatic performances, where the “films" test uses audio-visual recordings from movies that shows complex situations. These behavioral tests target the Theory of Mind (ToM), that is the ability to attribute mental states (beliefs, desires, intentions and emotions) to others that are distinct to one’s own. ToM being a crucial part of cognitive empathy, these simple tests allow to spot deficiencies while targeting necessary perceptual abilities.

Understanding appraisals (Baron-Cohen et al., 1986), (Lawson et al., 2004) and intuitive physics (Baron-Cohen, Wheelwright, Spong, et al., 2001) can also be used to determine the capacity to understand cause and effect relationships, which is based on the higher level cognitive mechanisms. The Picture-Stories task (Baron-Cohen et al., 1986) consists of a series of pictures that show the cause and effect relationship in social situations when appropriately sequenced. Similarly, the Social Stories Questionnaire (SSQ) (Lawson et al., 2004) consists of 10 short stories that may involve situations where one character could upset the other character in the story. Participants are asked to whether a selected utterance from the story contains an upsetting utterance and whether the behavior of one character could have upset the other character. The number of correct answers defines the SSQ score in this test.

4.4 Evaluation of Empathy in Interactive Agents

Being a novel field, empathy studies in artificial intelligence (AI) has no strong standardization and validated methods to measure empathy in artificial agents. In the previous section, we laid out some of the most accepted evaluation methods to evaluate empathy in humans. Although these methods are well-established and agreed upon in the academic community, applying them in the context of artificial agents is not straightforward. Most of these tools rely on self-measurement which cannot be applied to computational systems or the assessment of an expert that is difficult to automatize. Moreover, the differences in the capabilities of the agents and the application context restrict the usage of general behavioral measurements. These issues made it challenging to use this know-how to the evaluation of empathy in artificial interactive agents.

Empathy measurements in psychology literature include the evaluation of specific cognitive and behavioral capabilities as well as an overall evaluation of empathy. Specific features include evaluations of emotion recognition (Baron-Cohen, Wheelwright, Hill, et al., 2001), perspective taking (Davis et al., 1980) and empathic concern (Davis, 1983). Understanding the user's emotion depends on the correct recognition of the facial expressions and the performance of the emotion classifier. The perception of empathic behavior depends on the successful expression of the intended empathic emotion. Overall evaluation of empathy should take these feature's performance along with the system-level evaluation of empathy.

Similarly, the performance of computational systems highly depends on the performance and accuracy of the individual components as well as the integration of these components at the system-level. Due to the complexity and multi-component nature of interactive agents, scholars suggested (Dybkjaer, Bernsen, & Minker, 2004; Ruttkay, Dormann, & Noot, 2006) to provide feature-level and system-level evaluations separately. System-level evaluations focus on the behavior of the agent as a whole, where feature-level evaluations are aimed to isolate individual components of the system separately.

Following this notion, we propose to combine the best practices in the HCI research with the traditional methods of evaluating empathy. In the following sub-sections, we will focus on how to evaluate empathy using system-level and feature-level evaluation methods. We will systematically list the factors that contribute to the evaluation of empathy in artificial agents to initiate the discussion towards creating a common ground.

4.4.1 System-Level Evaluation

System-level evaluation is focused on the measurement of the behavior of the system in a broader sense. Similar to the self-report and perceived empathy evaluations that are aimed at capturing the global empathic behavior, system-level evaluations in artificial agents focus on the overall perception of empathy of the agent. In these type of evaluations, the participants interact with the complete system according to the interaction context, and a set of subjective and objective evaluations are used to compare the different versions of the system or with human behavior.

Previous studies in artificial empathy often focus on the second person or third person perception of empathy of the systems by using empathy-related terminology such as “feeling with" (Rodrigues et al., 2015), “feeling with" (Boukricha et al., 2013), “emotion matching" (McQuiggan et al., 2008), “compassionate" (Ochs et al., 2012) or “caring" (Brave et al., 2005). However, these terms only focus on one specific aspect of empathic behavior or related constructs. An interesting approach was used to train and evaluate the CARE framework (McQuiggan et al., 2008) by comparing the behavior of the agent with human behavior in a goal-directed environment. This approach can be automated but requires additional data-collection and evaluation steps of human behavior in a similar context to allow for a direct comparison.

As computational empathy research gaining more attention, researchers are beginning to raise awareness on the importance of using more suitable metrics. Recently scholars (Paiva et al., 2017) suggested using a variation of the IRI questionnaire (Davis, 1983) by adapting the first-person evaluation to a perceived-empathy survey. This idea was applied as a part of the EMOTE project (Barendregt, 2016) authors assessed the perceived empathy of a social robot using the IRI questionnaire. Similarly, Toronto empathy questionnaire (Spreng et al., 2009) was used as a perceived empathy metric by converting the self-report questionnaire into a second or third person evaluation (Yalçın & DiPaola, 2019). These evaluation methods can be used to evaluate the system by the interaction partner using a questionnaire. Moreover, the evaluation can be done by a third-person after watching the live or recorded interaction between the system and a participant. Although these methods provide an evaluation that is aligned with the related research on empathy, they were not validated.

However, these perceptual evaluations of empathy can be affected by several factors that should be taken into consideration while applying these system-level evaluations. These factors can be categorized as user-related factors, context-related factors and system-related factors.

User-related Factors

Research on empathy shown that humans empathize with each other on different levels depending on factors such as their gender, mood, personality, similarity and social capabilities (de Vignemont & Singer, 2006; Davis et al., 1980; Hoffman, 2001). These findings highlight the importance of controlling for these factors in a comparative evaluation of the agent behavior. Moreover, individual traits such as culture, socio-economic background and computer experience might affect the evaluation of the system as an interactive tool (Reeves & Nass, 1996).

Context-related Factors

Relationship and context related factors would impact the strength and expression of empathic behavior. The context, the appraisal of the situation or the social role of the empathizer are suggested to influence the regulation of emotions (Omdahl, 2014). Systems that act as companions as opposed to trainers are expected to be more friendly. This user expectancy based on the role of the agent and the context can effect the perception of empathy, where people tend to be more empathic towards in-group members such as friends and family members (Hoffman, 2001; de Vignemont & Singer, 2006). Moreover, goal-directed factors that show the quality of experience such as effectiveness, efficiency, user-satisfaction, utility and acceptability can influence the overall perception of the system (Moller, Engelbrecht, Kuhnel, Wechsung, & Weiss, 2009; Ruttkay et al., 2006).

System-related Factors

Factors related to the system behavior that are not directly linked to its empathic capacity can also impact the evaluation of empathy. Studies have shown that aesthetic characteristics of the interaction partner have a dramatic influence on the perception of empathy in humans (Müller, Van Leeuwen, Van Baaren, Bekkering, & Dijksterhuis, 2013). These aesthetic considerations might translate into the factors related to the looks, human-likeness, fluency of movement and believability of the agents (Misselhorn, 2009; Loyall, 1997). HCI research has developed evaluation metrics to control the effect of these factors such as anthropomorphism, animacy, likability, perceived intelligence and perceived safety (Bartneck, Kulić, Croft, & Zoghbi, 2009).

Computational empathy research has already been measuring some of these factors as control variables as well as additional metrics for the overall success of their system (Ochs et al., 2012; Rodrigues et al., 2015; Leite et al., 2014). Although the effects of the factors related to empathic behavior are examined in detail in empathy research, the relationship between these factors and the perception of empathy is yet to be examined.

4.4.2 Feature-Level Evaluation

In addition to the system-level evaluation, the evaluation of individual aspects of the system is necessary to assess the empathic capabilities of an interactive system. Feature-level evaluations can provide an incremental assessment of each component and capability of the agent. This allows for capturing the propagation of errors in empathic behavior, similar to the behavioral evaluations in empathy research that focuses on capturing deficits in empathic capacity.

In complex interactive systems, the evaluation methodologies usually include the metrics from various sub-fields, such as speech recognition, emotion recognition and speech synthesis (Ruttkay et al., 2006). The performance of the implementation is depended on the success of the separate features of the system, as each component affects the evaluation of other components. Therefore, the deficits in one capability might drastically influence the other. For example, the appraisal mechanism could be effected by simply a poorly performing emotion recognition component. Similarly, according to the empathic capabilities implemented to an agent, the features of every capability should be evaluated systematically at every stage of development.

For the evaluation of empathy as a broader concept, we will use the categorization of empathy features based on the evolutionary approaches (de Waal, 2007; de Waal & Preston, 2017) as we discussed in Section 4.3. According to these approaches, the empathic capacity can be categorized into three hierarchical mechanisms: emotional communication, emotion regulation and cognitive processes. Similar components have been proposed by other researchers in empathy (Paiva et al., 2017; Yalçın & DiPaola, 2019) and emotional intelligence research (Scherer, 2007). However, it should be noted that different types of definitions, models of empathy, as well as the capabilities and goals of interactive agents would require the evaluation of different subsets of these capabilities.

Emotional Communication

Emotional communication capacity forms the foundation of affective behaviors including empathy (Scherer, 2007; de Waal & Preston, 2017). This capacity can be further categorized as emotion recognition and emotion expression components. The successful detection and recognition of the input emotions would directly impact the empathic behavior of every level, hence the perception of empathy of the agent. Similarly, as the empathic behavior is essentially an emotional response to the stimuli, the emotional expression ability of the agent would directly influence the empathic behavior and the evaluation of the behavior. Therefore, it is crucial to include the individual evaluations of emotional communication capacity to assess the empathic capabilities of an interactive agent.

The evaluation of the emotion recognition ability may include a variety of well-established tests depending on the input modalities of the agent. For example, a text-based conversational agent's emotional communication capability can only be tested via the text-based linguistic emotional recognition and expression, where an embodied conversational agent should also be evaluated according to its speech, body gestures and facial expressions. Similarly, the success of the emotion expression behavior should be evaluated depending on the output modalities of the agent that are going to be used for expressing the empathic emotions. Following the behavioral metrics for empathy that are designed to evaluate the emotion recognition from pictures (Baron-Cohen, Wheelwright, Hill, et al., 2001), voice (Golan et al., 2007) and complex emotions from movies (Golan, Baron-Cohen, Hill, & Golan, 2006), the metrics for the recognition of agents should include the evaluation of each modality. Affective computing research provides well-established evaluation metrics for emotion recognition in computational systems (Scherer et al., 2010).

Emotion Regulation

A variety of models on empathy and emotional intelligence assign central importance in the ability to regulate emotions based on a variety of dynamics (Scherer, 2007; de Waal & Preston, 2017). Emotion regulation capacity can be based on personality and mood of the individual that allows for automatic regulation (Scherer et al., 2010). Humans are found to automatically assign attributes such as personality, gender and mood to interactive systems (Reeves & Nass, 1996). Personality metrics such as the Big Five are widely used in affective computing research (Vinciarelli & Mohammadi, 2014). Subjective evaluation metrics for emotional control have been proposed (Gullone & Taffe, 2012; Preece, Becerra, Robinson, Dandy, & Allan, 2018). However, these approaches have not been used by the empathic computing research and may require adjustments.

Cognitive Processes

The higher level of empathic capacity is suggested to include the cognitive processes such as appraisal, re-appraisal, self-oriented perspective taking and other-oriented perspective taking behaviors (de Waal & Preston, 2017). These cognitive processes would also control the emotion regulation abilities that allow for suppression or enhancement of emotions based on the context (Scherer et al., 2010). As we covered in Section 4.3, behavioral measures such as understanding appraisals (Baron-Cohen et al., 1986), the picture-stories task (Baron-Cohen et al., 1986) and the Social Stories Questionnaire (SSQ) (Lawson et al., 2004) are used to assess the deficiencies in the cognitive empathy. However, there are no standardized method to evaluate these capabilities in artificial agents. Moreover, the domain dependence and the problem of scalability for the cognitive capabilities makes it problematic to perform these tests to interactive agents with various capabilities.

Even though we suggested solutions for the feature-level evaluations to adopt the existing metrics, most of them needs further adjustments and validations to be applied in artificial agents.

4.5 Concluding Remarks

Empathy as a complex socio-emotional phenomena where the variety of definitions and models in the research community makes it problematic to evaluate and compare the implementation of the behavior in interactive artificial agents. This article is aimed to describe in detail the methods have been developed in the empathy research to evaluate empathic behavior, that can be translated into the emerging computational empathy research. We attempted to provide a systematic approach to the evaluation of this complex by suggesting the approach the evaluation on system-level and feature-level. As we acknowledge the difficulties of establishing a common ground in a diverse set application areas and capabilities of agents, we believe the importance of specifying the broader picture in the evaluation of empathy.

Our goal was to provide a guide on how we can evaluate the empathic behavior in artificial agents. We proposed system-level and feature-level evaluations for computational empathy systems to approach the issue systematically. We further provided a list of factors and components that can be used as a road-map to create individual evaluations for empathic systems in various application areas. We propose that the extensive body of work in the evaluation of empathy in humans, and the evaluation methods from affective and social computing can be used for computational empathy research. We hope to initiate the discussion towards creating a common ground to evaluate and compare computational empathy methods with this paper. We believe that a collective effort is required to develop specific measures and evaluation frameworks of empathy for interactive artificial agents.

Chapter 5

Empathic Listener Framework for Embodied Conversational Agents

This paper is accepted in Cognitive Systems Research Journal: Yalçin, Ö. N. (in press). **Empathic Listener Framework for Embodied Conversational Agents**. *Cognitive Systems Research Journal*, (pp. xx-xx). Contributions: I was the main and sole contributor to this paper.

In this paper, I present a computational framework of empathy for embodied conversational agents and provide details of the implementation of levels of empathic behavior. The framework focuses on the individual theoretical components presented in Chapter 3 and present implementations for the emotional communication and affect regulation levels. I also propose recommendations for the implementation of the cognitive mechanisms and how they can be integrated into the framework.

5.1 Abstract

Empathy is a complex socio-emotional behavior that results from the interaction between affective and cognitive mechanisms. Equipping embodied conversational agents (ECAs) with empathic capacity can benefit from the integration and evaluation of these low and high level capabilities in a hierarchical manner. Following the theoretical background on empathic behavior in humans, this paper presents a framework to equip ECAs with real time multi-modal empathic interaction capabilities. We present the implementation of this framework, which includes basic dialogue capabilities as well as three levels of empathic behavior in a conversational scenario. Our approach is an inclusive stand on modeling levels of empathy and provides a baseline behavior for empathic interaction.

5.2 Introduction

The ability to understand and react towards the emotions of others, empathy, is a crucial socio-emotional behavior for smooth interpersonal interactions. There has been an increasing amount of contributions to model empathy in agents with various types of behavioral capabilities(Paiva et al., 2017) as an attempt to enhance the interaction between humans and artificial agents and understanding the empathy mechanism better. Embodied conversational agents (ECAs) shows promise as an application area for artificial empathy due to their focus on the importance of natural multi-modal interaction.

However, modeling empathic capacity on interactive agents is a challenging task. Empathy can be attributed to a range of behavior from mirroring, affective matching to empathic concern, altruistic helping and perspective taking (de Waal & Preston, 2017; Coplan & Goldie, 2011). These low and high level behaviors result from the complex interaction of hierarchical components which can be clustered as emotional communication mechanisms, emotion regulation mechanisms and cognitive mechanisms (Yalçin, 2018). It is beneficial to understand how much each of these components contributes to the perception of empathy in interactive agents.

In this paper, we present a framework for creating empathy-enabled embodied conversational agents, a critical problem in ECAs. The framework leverages three hierarchical levels of capabilities to model empathy for ECAs, as suggested by the theoretical models (Yalçin, 2018; de Waal, 2007). This framework can be used as a building block for integrating additional empathic and behavioral capabilities while having an efficient and responsive conversational behavior. Our system incorporates these baseline empathic behaviors by equipping emotion communication capabilities, which are perceiving and expressing emotions. We present an empathic dialogue scenario to evaluate the implementation of our framework in isolation, focusing on the changing states of the dialogue during the conversation.

In the following sections, we examine the related work on empathy and embodied conversational agents. Then, we provide the interaction scenario and the proposed framework for our 3D virtual avatar system. We further explain how the information processing flows between different components can lead to the levels of empathic behavior.

5.3 Related Work

Computational Empathy is a relatively new research area that focuses on implementing empathy in virtual agents. Embodied conversational agent (ECA) research as an active and successful field has been attempting to integrate empathy to the existing ECA frameworks. However, while applying the know-how of this mature area to the novel space of computational empathy, it is important to pay attention to a possible overlap between the existing components of these frameworks with the requirements of empathy. A close examination of the theories of empathy would allow us to develop a framework that can account for this possibility.

Current research on empathy in embodied conversational agents (ECAs) follows a variety of different theoretical approaches to define and model empathic behavior (Paiva et al.,

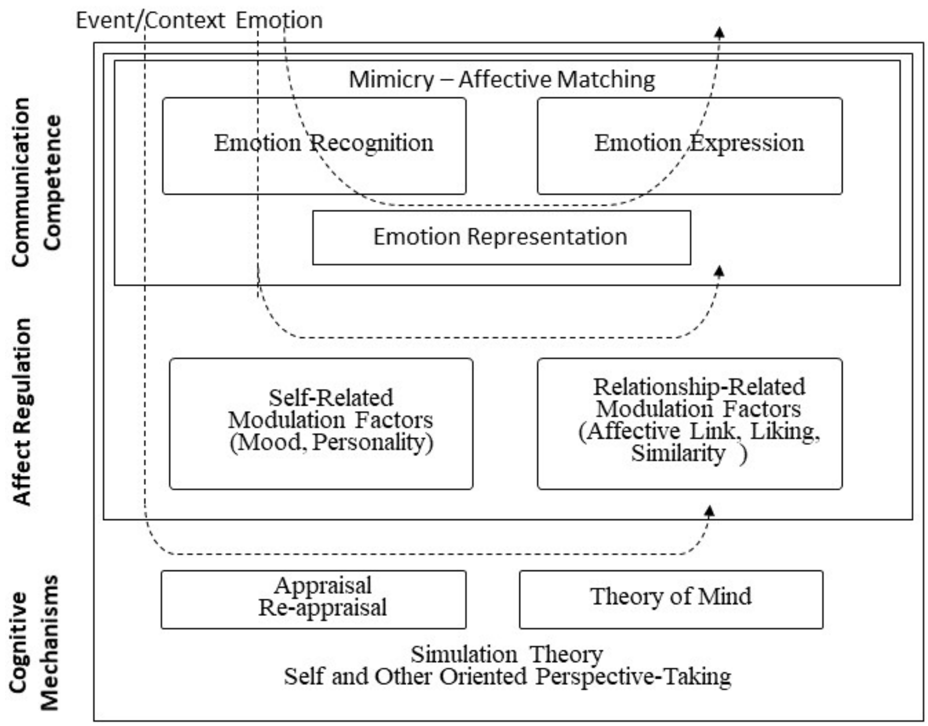

Figure 5.1: A model of empathy that involves hierarchical levels that are responsible of various types of empathic behavior.

2017). Most of the studies focus on a binary classification of empathy in artificial agents (Brave et al., 2005; Prendinger et al., 2005). Ochs and colleagues (Ochs et al., 2012) provide a formal model of emotions based on appraisals while concentrating mostly on the cognitive evaluation of emotions by the agent. Another approach by Rodrigues and colleagues (Rodrigues et al., 2015) incorporates emotion regulation components such as similarity, affective link, mood and personality that allows different intensities of emotion to be presented. Boukricha and colleagues (Boukricha et al., 2013) addressed the modulation of empathy according to the mood and familiarity of the agent and the perceived emotion. A third-person evaluation of empathy included three levels of empathy in a scale from feeling cold-towards and feeling-with in different familiarity and liking conditions.

An inclusive model of empathy would require the development of three hierarchical levels of empathic capabilities that can be categorized as communication competence, affect regulation and cognitive mechanisms (Yalçin, 2018). It is crucial to develop and evaluate the system components separately while being mindful about the resulting behavior that each component is responsible for, due to the complexity of the interaction of these components. Research on empathy from various disciplines suggests the underlying mechanism for any empathic behavior to be the perception, recognition and expression of affective stimuli (de Waal & Preston, 2017; Paiva et al., 2017). Figure 5.1 shows a model of empathy by Yalcin and DiPaola (Yalçin, 2018), which suggests a three-level hierarchy of empathic behavior inspired by the work of de Waal (de Waal & Preston, 2017). In this model, communication competence allows for mimicry and affect matching behaviors while forming a foundation to higher levels of empathy. Emotional mimicry is considered as the congruent affective response to the observed individual's emotions, either by matching the motor expressions or the emotional representation of it in the observer (Hess & Fischer, 2014). This capability is essential for any type of empathic behavior (de Waal, 2007) as well as emotional intelligence in general (Scherer et al., 2010).

Moreover, for an ECA to be perceived as a social agent that is capable of natural interaction, requires it to follow some basic principles of human conversation (Schroder et al., 2012) as well as showing emotions (J. Bates et al., 1994). As a socio-emotional behavior, it is expected that the perception of empathy also affected by this. Natural human interaction consists of non-verbal and verbal behavior (McNeill, 1992) that includes multi-modal synchronous behavior speech, intonation, gaze, head movements and gestures make meaning together using different channels at different timescales (Gratch et al., 2002). Listening behavior in humans consists of a combination of head nods, vocalizations and facial feedback that show agreement and acknowledgment, which is called backchannels (Maatman, Gratch, & Marsella, 2005). Turn-taking and backchanneling acts (Duncan, 1972; Charles, 1981), as well as the context of the dialogue, are what determines which of these will be used by the conversation partners while they are in a speaker or a listener role. However, the integration of these behaviors in an empathic ECA should be carefully examined as the output gestures and synchronization might intersect with the expression of empathic behaviors during an interaction.

The ECA literature provides a detailed examination of several types of backchannel behavior during the listening act, but offer little insight on empathy behavior. Sensitive Artificial Listener (SAL) (Schroder et al., 2012) is a multimodal dialogue system that is capable of nonverbal interaction based on speech, head movement and facial expressions of the user. This important work takes into account the user's emotion during listener feedback and can provide mimicry based on the detected facial action units (AUs) of the user. The system is also tested on different characters that had various personality factors that effect the selection and expression of emotional feedback. It was found that the emotionally congruent listening feedback results in better rapport and perception of social presence. However, empathy perception was not a part of this study and there was no distinction between the different types of affective feedback during evaluation.

Similarly, Skowron (Skowron, 2010) focuses on a specific type of conversational agents, which is called Affect Listeners that are capable of detecting affective responses, reasoning with them and responding. However, the system and the evaluation is based on task-specific scenarios that are aimed at increasing user satisfaction and usability only. The Rapport Agent of Gratch and colleagues (Gratch, Wang, Gerten, Fast, & Duffy, 2007) found that random feedback is worse than contingent feedback, where the frequency of feedback was constant. They found the mimicking of head nods according to the prosodic features were perceived as an increased emotional rapport compared to random feedback. This suggests that even the specific backchannel behavior which seemingly does not have an emotional value attached to it when observed in isolation can have an emotional effect on the perception of the user. This highlights the importance of equipping the agent with backchannel feedback while providing a comparison based on the perception of empathy. Previous research repeatedly shown that affect sensitive feedback improves the interaction (Ball & Breese, 2000; J. Bates et al., 1994; Brave & Nass, 2003). However, the literature does not give us insight into how the perception of empathy might be affected by this change.

In this paper, we propose a framework for embodied conversational agents (ECAs) that would allow us to implement levels of empathic capacity along with basic conversational behaviors. We aim to provide a hierarchical implementation of each empathy level along with controlled integration of conversational capabilities, to be able to test and compare each level and component with each other. This paper provides the framework with the current implementation for an empathic listening agent and explains how levels of empathic behavior can arise from different information processing cycles. In the following sections, we will present an empathy framework for our ECA starting with the description of the interaction scenario.

5.4 Agent and the Context

Our framework is intended for an embodied agent capable of maintaining conversational interaction with its human counterpart. This requires certain behavioral capabilities, such as being able to receive conversational input, process it and act in a social and emotionally suitable manner. In this context, our framework is implemented in a human-like embodied conversational agent. Our socially-situated 3D virtual character system can perform a set of verbal and non-verbal behavior that allows for a realistic conversation with the interaction partner. These behaviors include facial expressions, gaze, head and body gestures, as well as verbal behaviors.

The interaction scenario includes a face-to-face conversation between the agent and a human interaction partner, similar to a video-conference. According to the state of the dialogue, the behavior of the agent can change and adapt to the user. While the interaction partner is speaking, the agent enters the listening state. Listening mode will be activated via the speech and video input from the agent that was described in the perceptual module section. In this state, the agent is expected to provide proper backchanneling behavior as well as the emotional feedback. After the speech of the interaction partner is completed, the agent will enter the thinking state. In this state, the agent will be finished gathering information from the perceptual model and start processing the speech input for generating a response. This response generation process will make use of the context of the dialogue as well as the emotional content of the message. Lastly, the agent will enter the speaking state, where it executes the prepared response via its output channels including voice, facial expression and body gestures.

Note that, our approach includes discrete stages of the dialogue, where each stage sequentially follows each other. Nevertheless, thinking, speaking, and idle states of the agent can be interrupted by the user, which allows for more complex interaction. Moreover, the same framework can be used for agents equipped with different output channels, as long as it allows for a mapping between the emotion representations and the behavioral capabilities. Next, we will go over how the agent framework incorporates levels of empathic capacity within this conversational interaction scenario.

5.5 Empathy Framework

Our framework for an empathic agent is intended to implement an ECA that is capable of responding to an emotional conversation with the user using verbal and non-verbal behaviors. The implementation achieves levels of empathic behavior through the interaction between the components of perceptual and behavior generation modules with the central behavior controller. The perceptual module processes the visual and verbal input from the user to analyze the context and emotional value. This information is then sent to the behavior controller to be reasoned with according to the selected empathy mechanisms. According to the level of empathic behavior and the context, an emotional response is selected and sent to the behavior generation module to prepare the response to be displayed with the behavior realizer. Figure 5.2 shows the overall framework of our system.

Levels of empathic behavior are achieved by following various processing pathways within this framework. Low-level empathic behavior such as mimicry and affective matching goes through minimal processing in the behavior controller, which allows for a fast response. Mid-level empathic behavior such as affect regulation requires handling the dynamic properties of emotions that can change according to temporal parameters such as mood and personality. Lastly, the higher-level empathic behavior includes reasoning with information about the context. Due to the use of inputs from the lower processing levels, the high-level reasoning results in a slower response.

The following sections will provide detailed information about the modules of the framework and its implementation. The framework is implemented using Python 3.5 programming language, and the output of the framework is realized using Smartbody behavior realizer in Windows platform [1]. Inputs are gathered using a standard web-cam and a microphone.

[1]The implementation code and documentation is released at https://github.com/onyalcin

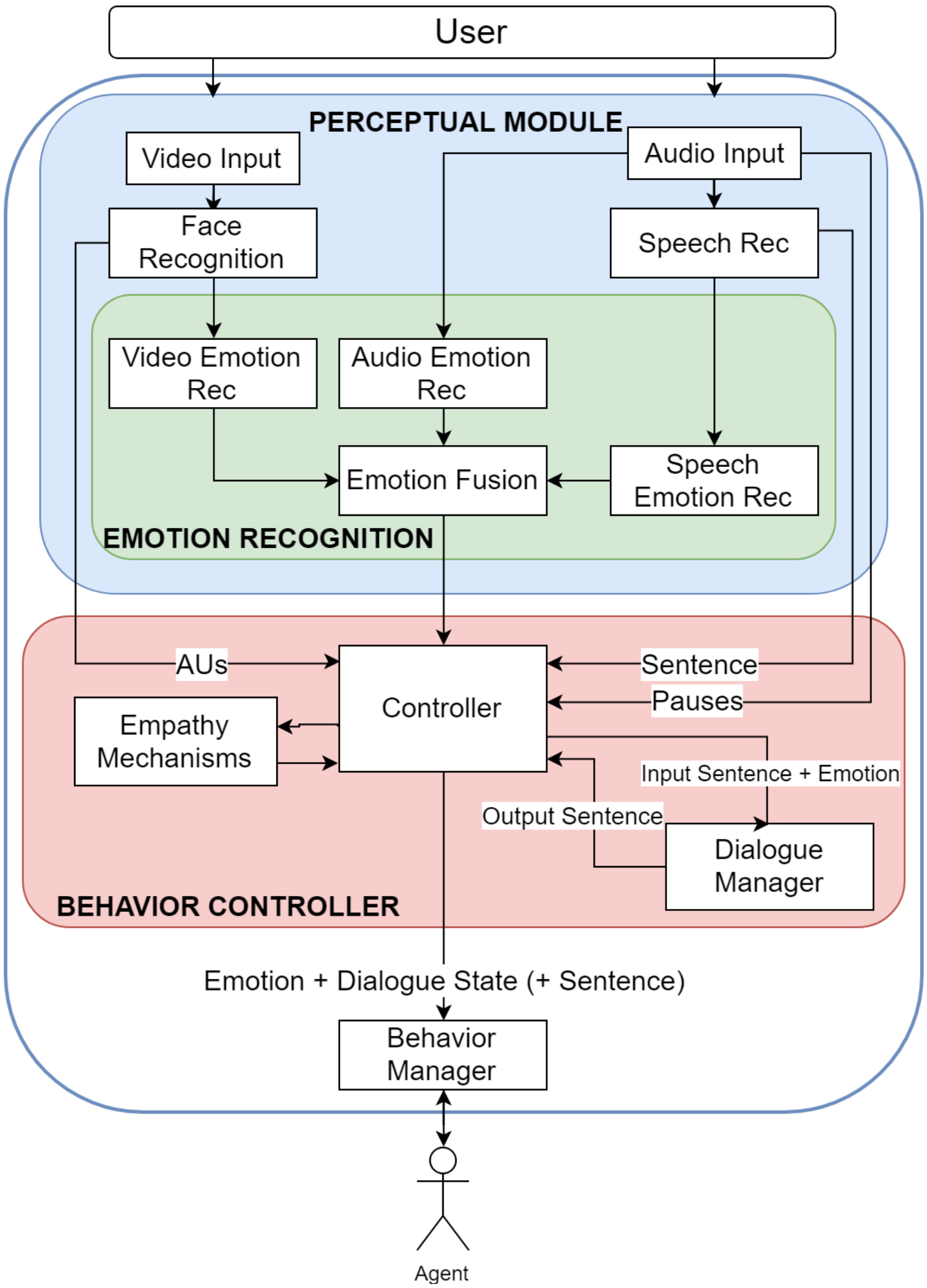

Figure 5.2: Our framework that includes perceptual, behavior controller and behavior manager modules. Perceptual module includes a multi-modal emotion recognition sub-module. Empathy mechanisms are selected via the controller and impact the behavior of the virtual agent accordingly.

5.5.1 Perceptual Module

The perceptual module manages the visual and audio inputs received from the user and generates representations of these inputs to be used by the behavior controller. These rep-

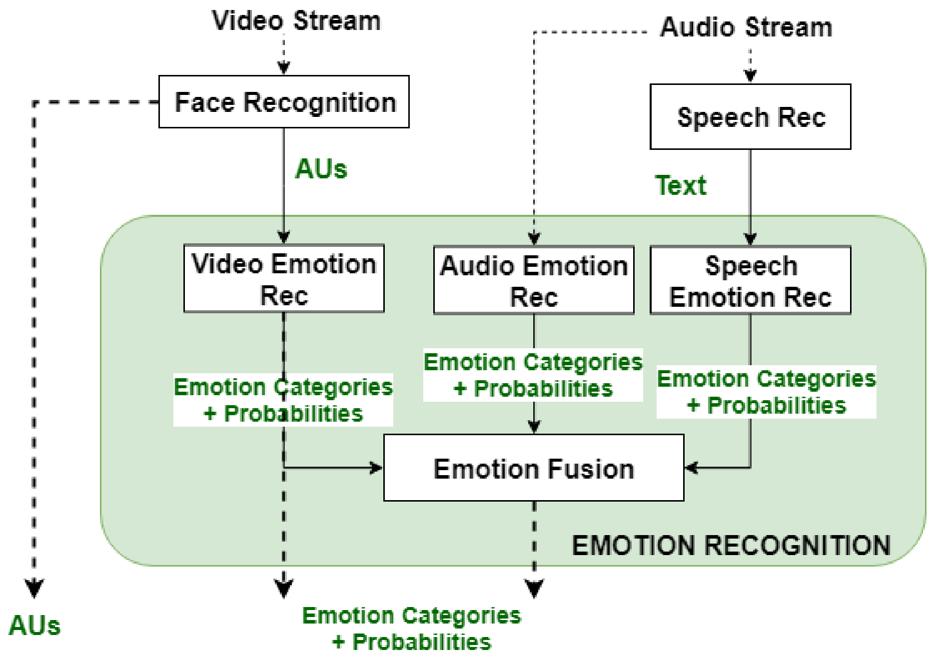

Figure 5.3: Perceptual module processes video and audio inputs received from the user on separate streams. The outputs from this module are sent to the Behavior Controller according to the state of the agent and the level of selected empathic behavior.

resentations can be low-level such as the changes in the energy levels from the user's voice based audio signal. Higher-level representations such as emotion categories from the voice and facial emotion recognition, as well as the words and sentence-level sentiment are also gathered in this module. Figure 5.3 shows a detailed graph of the processing streams within this module.

Audio input includes verbal signals from the user, where the initiation, pauses and termination of the speech signal are gathered to inform the behavior controller about the state of the conversation. This information is gathered using the energy levels from the audio signal with PyAudioAnalysis library (Giannakopoulos, 2015), and an active speech channel is recognized via a Push-to-Talk functionality that we developed. Speech is recognized by using the real-time Text-to-Speech (TTS) system by using Python's SpeechRecognition (A. Zhang, 2017) library with Google Cloud Speech API for Python (Google, 2014-2017). The streaming speech signal and the pauses within the speech signal are sent directly to the Controller during listening to be used to create backchannel behavior, such as head nods and tilts, in real-time. The audio and the text output are also sent to the emotion recognition sub-module for further processing.

Video input is used to extract facial landmark points to be used in the emotion recognition system. We currently use Affdex real-time expression recognition toolkit (McDuff et al., 2016) to extract facial expressions for emotion recognition. During the interaction, the perceptual module collects information about the facial action units (AUs) that include AU1 (inner brow raiser), AU2 (outer brow raiser), AU4 (brow loweverer), AU5 (upper lid raiser), AU6 (cheek raiser), AU7 (lid tightener), AU9 (nose wrinkler), AU10 (upper lip raiser), AU12 (lip corner puller), AU14 (dimpler), AU15 (lip corner depressor), AU17 (chin raiser),AU18 (lip puckerer), AU20 (lip stretcher), AU24 (lip pressor), AU26 (jaw drop), AU28 (lip suck), and AU43 (eyes closed). These are used both directly during the mimicry process and indirectly during affective matching while first passed through the emotion recognition sub-module before being sent to the Controller.

Emotion Recognition

The emotion recognition module is a component of the perceptual module that manages the recognition and fusion processes using the information gathered from the user inputs. Our implementation includes three types of modalities for the purposes of emotion analysis: facial emotion recognition from video, tone analysis from low-level audio information, speech emotion recognition from linguistic information. These sub-components are activated according to the selected empathy mechanism as well as the state of the dialogue. During listening, emotion recognition is not activated if the selected empathy mechanism is low-level mimicry behavior. For affective matching behavior during the listening state, the agent uses the immediate information from the video and audio emotion recognition sub-components. After the user is finished speaking, the agent enters the thinking state, where the overall emotion of the interaction partner is calculated by including the emotion recognition from the speech signal and sent to the controller for a proper response.

Video emotion recognition component uses OpenCV library (Bradski, 2000) for face detection and the CK Dataset (Lucey et al., 2010) for training. The face images are categorized in basic emotion categories (Anger, Disgust, Fear, Joy, Sadness, Surprise and Contempt) as well as Valence value in a weighted scoring system [2]. Emotions are recognized based on frames in 60 frames per second. Audio emotion recognition component is intended to use the audio signal from the audio input stream to detect the emotions from the speech signal. We used RAVDESS audio dataset (Livingstone & Russo, 2018) to train our CNN-based emotion recognition model[3]. The speech emotion recognition component is activated after the speech recognition component fully processes the speech signal, which explained in the previous section. The recognized speech is further processed in this component using the SO-CAL sentiment analyzer (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011) and NRC-Canada System (Mohammad, Kiritchenko, & Zhu, 2013).

[2]The facial emotion recognition model can be found in https://github.com/onyalcin/face_emotion_-recognition

[3]The audio emotion recognition model can be found in https://github.com/onyalcin/audio_emotion_-recognition

The outputs from these components are passed to the emotion fusion component along with the output from audio emotion recognition. We aim to use late-fusion classifier for combining the outputs of the different modalities to detect the emotion of the user during the user's speech. This process is done by providing a weighted scoring system for each basic emotion category: anger, disgust, fear, joy, sadness, surprise and contempt.

5.5.2 Behavior Controller

The central component of this framework is the behavior controller, which acts as a decision-maker that provides a link between perceptions and actions. Depending on the state of the conversation and the empathy mechanisms to be used, the behavior controller selects the input channel to be used, processed and evaluated for preparing a response. The decision to react to the emotional input from the user emphatically is made in this central module, where the emotion of the agent itself is regulated based on the percepts. During a conversation, the agent should be able to decide the conversational state depending on the interaction: listening, thinking, speaking and idle. Furthermore, at each one of these states, the behavior of the agent should change according to the emotional value (valence, arousal or the emotion category) as well as the empathic behavior the agent is assigned to.

Behavior Controller includes three sub-modules: Central Controller, Empathy Mechanism, and the Dialogue Manager. These sub-modules process the information received from the perceptual module and generate a response using different strategies according to the state of the dialogue the agent is in. The Controller is responsible for the timing and sequence of processing for each sub-module. It determines which channels to receive information from, how to process them, in what order to process them, and how to generate the response.

During the listening state, the agent uses the pauses during the user's speech to give backchanneling behavior and the facial emotion recognition results to provide a matching emotional gesture (see Figure 5.4). After the speech of the user is over, the agent will enter the thinking state while processing the message within the Behavior Controller according to the selected empathy mechanism (see Figure 5.5).

Language has an essential role in the perception and experience of emotions and has been suggested to be equally important for empathy(Omdahl, 2014). Therefore, empathy mechanisms are highly linked with the dialogue manager in the behavior controller module. The Dialogue Manager is responsible for generating proper verbal responses according to the utterance of the interaction partner, as well as the emotional value of the desired response that is created by the Empathy Mechanisms sub-module. The Empathy Mechanisms sub-module embodies all the necessary emotion processing functions for generating levels of empathic behavior. The sequence of the processing steps changes according to the conversation state and the selected level within the empathy mechanism. Following the theoretical background, our framework includes three levels.

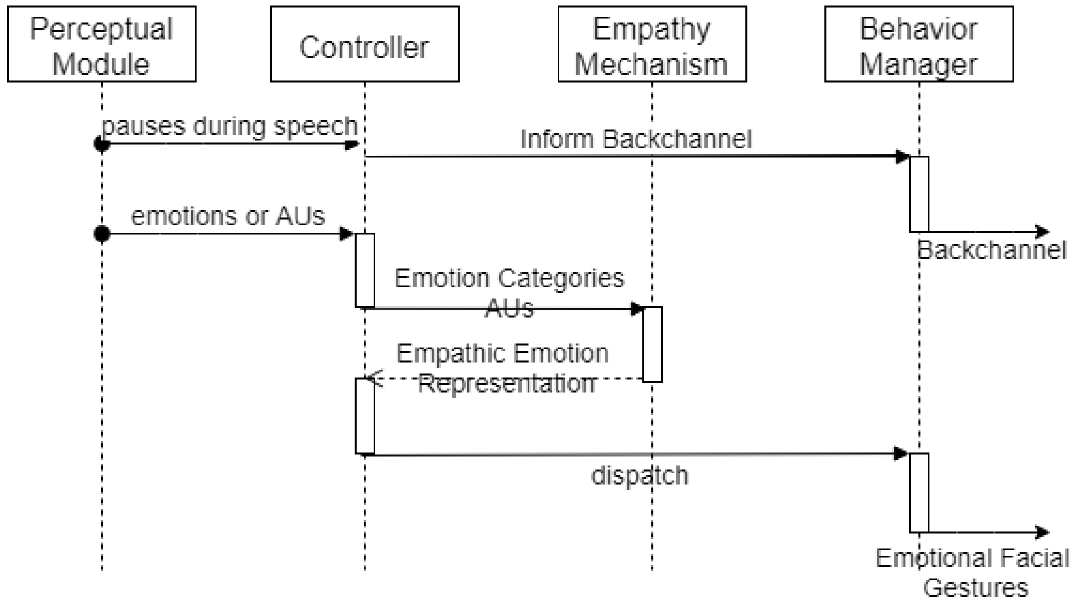

Figure 5.4: The information flow within the Behavior Controller during the listening state. It processes the pause signal from the speech and the emotion categories or AUs to give facial emotional feedback.

Low-Level Empathy

Low-level empathic behavior includes mimicry and affect-matching behavior (Yalçin, 2018), which allows for fast response mechanisms while relying on the Perception-Action Cycle (de Waal & Preston, 2017). This level of behavior requires less processing from the higher levels of cognitive functions and is achieved by either matching the perceived AUs or the perceived emotions with agent's expressive repertoire. During the listening, the Controller gathers the AU or emotion representation input from the perceptual module and sends it to the Empathy Mechanisms sub-module for matching representations in real-time. This information can be used for generating a matching speech response in the Dialogue Manager module (see Figure5.5). After the response is ready, the Controller again sends the emotional response to the Behavior Generation module for generating an output.

First evaluations of these capabilities showed that the implementation of our framework was perceived significantly more empathetic than baseline backchanneling behavior during the listening, where the difference between mimicry and affect matching is significant during speaking state of the agent (Yalçın & DiPaola, 2019).

Affect Regulation

Affect regulation abilities form the middle layer of the empathic behavior, where the agent regulates the dynamic properties of the emotions according to self-related and relationship-related parameters (Yalçin, 2018; Davis, 1983). These can include mood, personality, liking

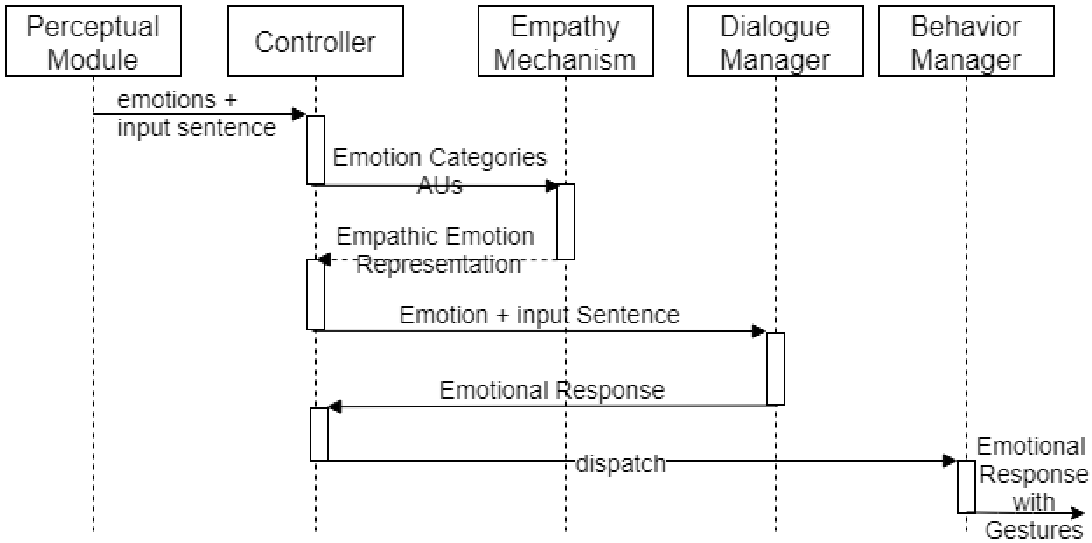

Figure 5.5: The information flow withing the Behavior Controller during the thinking and speaking cycles in low-level empathy behavior. The emotion of the user is first being processed to get a matching emotional behavior and sent with the input sentence to the dialogue manager to generate a response. This emotional response is then dispatched to the behavior manager to prepare synchronized behavior for the embodied agent.

and similarity parameters. As a part of the hierarchical process, regulation mechanisms would share the emotional representations generated by the lower-level functions (see Figure 5.6).

In the empathy mechanisms sub-module, mood and personality parameters changes the dynamic properties of the emotions in various time-scales. Mood allows the emotions to be sustained and decay, allowing for consistent behavior over time. The personality of the agent can change the frequency, amplitude and duration of the generated gestures. The mood of the agent can change during an interaction, while the personality parameters are not likely to change. Relationship-related parameters such as liking and similarity can also affect the strength of empathic behavior expressed by the agent. For example, pre-determined liking and familiarity values can be used to compute the similarity of the agent is expressed emotion compared to the emotion of the user (Boukricha et al., 2013; Rodrigues et al., 2015).

Cognitive Mechanisms

At the highest level of the empathy hierarchy, we have goal-oriented behavior that takes the context information into account (Yalçin, 2018). The Empathy Mechanism sub-module must include information about the goals of the agent, the event and the accompanying emotion in order to calculate the appraisal of the situation. The low-level information about the emotions will be received from the lower levels of the empathic mechanisms as well as

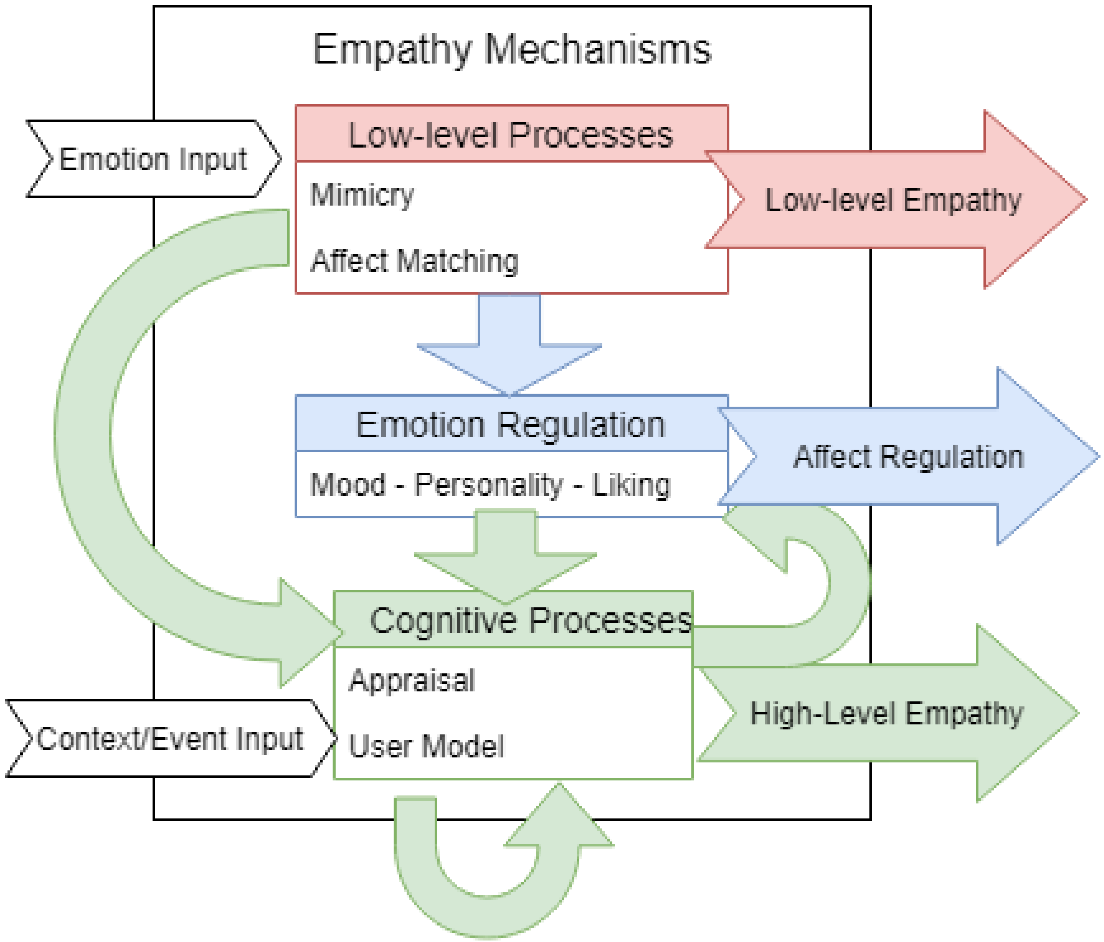

Figure 5.6: The process within Empathy Mechanisms sub-module allows for the generation of levels-of empathic behavior. The red colored arrows show low-level empathy flow, and blue arrows show the mid-level empathic behavior with affect regulation. Green arrows show multiple ways the higher-level empathy can be processed within the sub-module. The cyclic nature of information flow can be used for appraisal and re-appraisal of the situation with self and other-oriented perspectives.

emotion representations via the perceptual system. Information about relation from affect regulation level is also be used to compute the Information about context will be generated from the conversation; hence, the Dialogue Manager. This suggests that, unlike the lower levels of empathic processing, cognitive mechanisms require the processing of contextual information before being processed by the empathy mechanisms sub-component. Moreover, this cycle can be re-processed multiple times to allow for appraisal, re-appraisal and higher level recursive reasoning. This repetitiveness in the processing would result in a longer thinking state during the conversational cycle (see Figure 5.6).

The implementation of the appraisal processes can follow different theoretical models (Lazarus, 1991b; Scherer, 2001; C. A. Smith & Kirby, 2000; Roseman, 1996). In their paper, Ochs and colleagues (Ochs et al., 2012) provided a formal model of empathy according to the appraisal of the situation. Similarly, Broekens and colleagues (Broekens et al., 2008) suggested a formal model for Scherer's appraisal theory (Scherer, 2001) to use the representations of inputs from percepts to assign appraisal values based on some evaluation functions. However, an important distinction in cognitive processes related to empathic behavior is to model the other-focused appraisal of the situation, not the self-focused appraisal. A self-focused appraisal is the evaluation of the situation according to one's own goals and beliefs, where other-focused appraisal is the evaluation of the situation from the other's point of view. In that sense, a user model that relates to the Theory of Mind is a must. A good example of other-focused appraisal can be seen in the work of Lisetti and colleagues (Lisetti et al., 2013) that adopts user-modeling techniques for their goal-oriented embodied conversational agent.

5.5.3 Behavior Generation

The Behavior Generation module is responsible for generating the verbal and non-verbal response with proper timing and synchronization. It receives the emotion and proper output channels from the Behavior Controller module and prepares it for the agent interface, which is the Smartbody behavior realizer (Thiebaux et al., 2008). Smartbody allows for transforming messages received from the module into locomotion, object manipulation, lip syncing, gazing and nonverbal behaviors of the 3D agent in real time. We use the Stomp library (Briggs, 2018) to provide a two-way communication between the framework and the behavior realizer with Behavior Markup Language (BML) (Kopp et al., 2006). BML is a powerful standard in ECAs that allows for synchronization between multiple behaviors. Using BML standard, we can synchronize the exact timing of each gesture (body gestures, gaze, head movements, facial gestures) in relation to the speech of the agent.

The behavior generation component allows for synchronization of reflective behavior such as shifting body posture, breathing gaze or self-touching behavior as responses to the events happening in the environment and idle movement patterns during the lack of external input increases the quality of the interaction (Bernardet, Kang, Feng, DiPaola, & Shapiro, 2017; Bernardet & DiPaola, 2015; Nixon, DiPaola, & Bernardet, 2018). The consistency and coordination of these movements allow generating a sense of personality in affective agents (Bernardet, Saberi, & DiPaola, 2016).

5.6 Conclusion and Future Work

We have presented a framework and implementation of the levels of empathy behavior in embodied conversational agents following the theoretical background on empathy as well as the best practices in ECA research. We showed how each level in the hierarchical model of empathy could be realized via the interaction between different components of the framework in a conversational setting. This framework and implementation are intended to provide a blueprint for a wide range of applications with variations in input and output channels for emotion. Emotional representations linked to other bio-signals such as gaze behavior, breathing and heart rate, as well as non-anthropomorphic expressive behavior can be used following the layers of empathic processing.

5.7 Acknowledgements

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2019-06767] and the Social Sciences and Humanities Research Council of Canada (SSHRC) [435-2017-0625].

Chapter 6

Automated Affective Gesture Generation for Embodied Conversational Agents

This paper is submitted to the Journal on Multimodal User Interfaces: Yalçin, Ö. N., & DiPaola, S. (submitted). **Automated Affective Gesture Generation for Embodied Conversational Agents**. *Journal on Multimodal User Interfaces*, (pp. xx-xx).

Contributions: I am the main contributor to this paper. I was responsible for the conceptualization and implementation of the system as well as writing the original draft. Prof. DiPaola was supervising and reviewing the paper.

This paper presents the implementation details of the gesture generation process of our embodied conversational agent. The generation process is controlled by the emotion regulation processes. These processes are controlled by the parameters that determine the behavior of the agent in terms of expressivity, degree of variability and time-dependency. Implementation of this component ensures the automated generation of emotional facial and bodily gestures based on the long and short term dynamics of emotions.

6.1 Abstract

The growing success of dialogue systems research makes conversational agents a perfect candidate for becoming a standard in human computer interaction. The naturalness of communicative acts allows for providing a comfortable ground for the users to interact with. There have been many advances on using multiple communication channels in a dialogue system in the way of simulating humaneness in an artificial agent. However, engineering the usability of the system might have contradicting priorities compared with modeling a believable and natural agent. One issue is to be able to find a good balance of the intensity and frequency of multimodal affective feedback to guide the dialogue flow. The dynamics of the emotional feedbacks can have varying effects on different users. In this work, we propose a system for automating the generation of gestures according to the semantic and

affective properties of the response in embodied conversational agents. Our system is aimed to convey the intended affective message using multiple modalities.

6.2 Introduction

Emotions act as a facilitator for adapting to a complex environment and have an important role in the daily communications of humans (Scherer et al., 2010). Expression of emotions during conversation is achieved by using multiple channels of behavior such as facial expressions, head movements, body gestures, posture and gaze that accompany the verbal information (Russell, Bachorowski, & Fernández-Dols, 2003; Dael, Mortillaro, & Scherer, 2012; Kleinsmith & Bianchi-Berthouze, 2013; Adams Jr & Kleck, 2003). The synchronization and combination of these behaviors with the verbal utterances can fluctuate in time, due to the regulation of emotions. Understanding and expressing this dynamic interaction between the verbal information and the emotional gestures that accompany them are essential for successful and natural communication.

Embodied Conversational Agents (ECAs), are systems that are aimed to create a natural environment to interact with humans to achieve certain goals of the interaction. These agents can be used as health coaches (Bickmore & Picard, 2005), training systems (Serón, Baldassarri, & Cerezo, 2006), games (Becker, Prendinger, Ishizuka, & Wachsmuth, 2005), virtual assistants (Matsuyama et al., 2016) as well as tools for evaluating and examining human behavior (Kapoor, Burleson, & Picard, 2007). The premise of ECAs is that generating human-like behaviors in virtual agents would allow for an effortless interaction, due to the familiarity of these behaviors. ECA research showed that the expression of affective behavior has a positive effect on both human-human and human-agent interactions (Beale & Creed, 2009). Successful natural communication between an artificial agent and a human requires perceiving, understanding and responding to the emotional signals (Picard et al., 1995), which is also referred to as emotional competence (Scherer et al., 2010).

However, generating natural behaviors that accompany verbal behavior is a challenging task due to this dynamic and multi-modal nature of the speech gestures. Finding the right balance between expressivity and fluency is an important issue in real-time systems that can procedurally generate gestures that are appropriately synchronized with speech. While dialogue systems of ECAs are starting to be capable of producing complex emotional utterances with a minimum requirement for hand-crafting, gesture animation systems struggle to keep up with the flexibility needed to match these systems. Moreover, the dynamics of emotional behaviors depend on the intrinsic properties of the agent such as attitude, expressivity, personality and mood, as much as it is affected by the events in the environment. This dual interaction results in emotional expressions that vary in spatial and temporal characteristics which poses a need for an automated gesture generation system that is generalizable and parameterized.

This paper describes our implementation for the automatic generation of affective non-verbal behaviors that are naturally synchronized with the utterances of ECAs. Our focus is on providing consistent behavior depending on various emotional dynamics that allows for the regulation of emotions, namely the expressivity, degree of variability and time-dependency which are variables that help to provide a sense of attitude, mood and personality in agents. Our system is intended for the creation and use of virtual agents that can engage in a conversation by using bodily and facial movements synchronized with verbal expressions to achieve a natural interaction with a human user. Section 6.4 will provide an overview of the state-of-the-art in automated gesture generation in ECAs. The remainder of the paper will demonstrate our approach of automating the generation of gestures with emotional dynamics in detail.

6.3 Co-Speech Gestures

During conversation, humans use multiple channels to convey the meaning of their message that accompanies the speech. These channels include head gestures, eye gaze, facial gestures, bodily gestures and posture changes. The synchronization of these non-verbal behaviors in relation to the speech has been extensively studied by psychology and linguistics researchers. The categorization of gestures based on its spatial and temporal charactheristics is found to be essential to understand the synchronization of gestures with speech (Kendon, 1994). Kendon (Kendon, 1972) proposed an annotation scheme for the basic building blocks or "units" of gesture movements that consist of gesture phases: preparation, hold, stroke and retraction. The main phase is the "stroke" of the gestures that signifies the peak motion of the gesture that conveys the meaningful part. Holds can occur in between any of these phases that shows holding of the position. The preparation and retraction phases show the pre and post timing from the stroke of the gesture.

McNeil (McNeill, 1992) categorized gestures that accompany speech as iconic, deitic, metaphoric and beat gestures. Iconic gestures are gestures that have a physical relation between the object or actions during the speech. Deitic gestures are mostly pointing gestures that convey the spatial relations within the verbal information. Metaphoric gestures relate to the abstract contents of the speech, namely metaphors. Beat gestures, or batons in Ekman and Friesen's terms (Ekman & Friesen, 1972), synchronizes with the rhytmic components of the speech, mostly the prosodic features of it. McNeil (McNeill, 2006) also proposed the phases or the anatomy of these gestures in terms of its spatial and temporal characteristics to be able to annotate the synchronization of gestures according to speech. He proposed four basic stages, similar to Kendon's notion of gesture phases, pre-preparation position, pre-stroke hold, stroke, retraction and post-stroke hold. Only the "stroke" of the gesture is mandatory in these phases, where it signifies the meaningful component of the gesture. For an extended review of the gesture and speech interaction, see (Wagner et al., 2014) which gives an overview of gesture form and function in relation to parameters of speech.

These properties of gestures and their synchronization in relation to speech effect how the emotional information is conveyed during the speaking behavior. Most of the emotional information can be expressed using facial expressions (Ekman, 1993). These can be represented into discrete categories with the activation of groups of facial muscles that are called facial action units (FACS or AUs) (Ekman, 1972). Although most of the emotional information can be expressed using facial gestures, other behaviors also contribute to the emotional response such as gaze behavior, bodily gestures, posture and para-linguistic verbal utterances (Kleinsmith & Bianchi-Berthouze, 2013; Ekman, 2004). For example, "sadness" can be recognized via someone's facial gestures, head orientation, walking patterns as well as the tone and rhythm of their voice.

It was also argued that the emotional expressions are subject to change according to the personality traits, attitudes and regulation mechanisms, that contributes to a rich and natural interaction (Picard et al., 1995). An important variable that determines how many of these channels are used with what intensity is the expressivity, which is connected to the notion of personality and style. These include movement activity, spatial expansion as well as the power of the movement (Wallbott, 1998). As an example, extroverted gestures can be expansive and rapid while introverted gestures would be closer to the body and less in frequency (Gallaher, 1992).

Moreover, as words can often be representations of emotions, the emotional expressions rely on more than just the speech utterance (Ekman, 2004). The characteristics of the emotional expressions are also determined via dynamic parameters of the intrinsic properties of the agent. Emotional events are often responses to external events but are also subject to change according to the internal regulations (Kuppens & Verduyn, 2017). These two opposing forces determine how the agent will express its emotions in its behavior. Two fundamental variables that determine the emotional dynamics have been proposed as the degree of variability and time-dependency (Kuppens & Verduyn, 2015). The degree of variability determines the deviation from the mean behavior over time, where time dependency shows how much a gesture sustains over time. In our work, we make use of these parameters in order to generate a synchronized emotional behavior that matches the utterance as well as the intrinsic characteristics of the agent.

6.4 Related Work on Automated Gesture Generation

The ECA community have been exploring the automated generation of the gestures according to various behavioral parameters. Some researchers follow rule-based gesture generation systems, while others use data-driven systems that collect human movement data and annotate them according to gesture categories.

EMOTE system (Chi, Costa, Zhao, & Badler, 2000) implements the Laban movement model of effort and shape components in order to refer to the dynamic properties of behaviors. It uses expressivity, flow, weight, space and timing properties of the movement to generate a non-verbal emotional expression. One well-known example of an automated generation of co-verbal gestures is the BEAT framework (Cassell et al., 2004). BEAT is designed to generate gestures that are synchronized with the contextual and linguistic properties of the speech such as rheme, theme, word newness, contrast, objects and actions. They use iconic, contrast and beat gestures as well as eyebrow and gaze movements. This categorization is also present in our work which forms a baseline of gestures according to the utterance.

Other systems that are aimed to generate affect-related gestures use different behavioral parameters and focus separately on channels of expression such as facial gestures, body gestures, eye gaze and posture. Pelachaud and colleagues (Pelachaud, 2015; Poggi, Pelachaud, de Rosis, Carofiglio, & De Carolis, 2005) developed GRETA to model and execute complex expressions based on perceptual studies. Their system includes a gesture generation component to model and executes complex expressions based on perceptual studies(Niewiadomski, Bevacqua, Mancini, & Pelachaud, 2009; Poggi et al., 2005). It uses attack, delay, sustain and release parameters as well as spatial and power factors to express different emotions as gestures. Our system also uses behavior expressivity, spatial and temporal characteristics of gestures while focusing more on parameters of emotional dynamics and the distribution of gestures over the affective properties of speech.

Lee and Marsella (Lee & Marsella, 2006) created a non-verbal behavior generation system that analyzes the affective state, syntactic and semantic structure of the text to annotate the accompanying gestures. They used a data-driven approach to analyze the human-human conversation data using the SAL system (Schroder et al., 2012). The gestures they used were head, eyebrow and eye/gaze movements along with specific gestures such as shoulder shrug and labeled them according to the conversational acts such as affirmation, negation, intensification, contrast, response request, inclusivity, obligation and assumption.

Kipp and colleagues (Kipp, Neff, Kipp, & Albrecht, 2007) proposed a data-driven approach to non-verbal gesture synthesis while paying specific attention to the timing of gestures. They use Kendon's notion of gesture phases (Kendon, 1997) to automatically extract gesture data from human speakers and proposed an annotation scheme (Kipp, Neff, & Albrecht, 2007) to capture the temporal and spatial structure of a gesture. They identified basic movement phases as preparation, hold, stroke and retraction. The spatial dimensions are categorized according to the height, distance and radial orientation of the gesture in relation to the body. Our system uses a similar annotation scheme to evaluate the expressivity of the agent and to control the spatial and temporal properties of the gestures.

A data driven model by Chiu and Marsella (Chiu & Marsella, 2011) used a dataset of actors with simultaneous voice and gesture recordings in order to train a model that represents the relation between prosody and bodily movements. The model uses Conditional Restricted Boltzmann Machine that trained on pre-processed data, which includes upper-body movement sequences that are coupled with the pitch and intensity of the voice recordings for these segments. The semantics were not used in the data.

Different from these approaches, our system targets the intrinsic dynamics of emotion in combination with the external influence of the provided utterance. We make use of the meaning and affective value of the utterance. Furthermore, we pay attention to the emotional expressivity of our agent and do not rely on general rules to procedurally generate real-time affective gestures that are synchronized with the utterances.

6.5 Our Automated System

This paper shows the implementation of an affective behavior generator for coordination of gestures with the speech of ECAs in real-time. Our system is capable of selection and synchronization of gestures that reflects both the meaning of the speech and the desired affective meaning to be presented by the agent. We focus on the emotion dynamics and expressiveness of the agent to generate a consistent and natural behavior synchronized with the speech of the agent. Our system uses three distinct modules that are responsible for different types of processing, which will be covered in this section.

The input and output interaction of the system is done by a pipeline, which inputs the speech text and outputs a message containing the synchronized behavior prepared by the system to the behavior realizer. The output of our system for the generation of synchronized multi-modal behaviors such as speech, gaze, breathing, facial gestures, body posture and body gestures are realized in a 3D character in the Smartbody behavior realizer (Thiebaux et al., 2008). A two-way connection between our generation system and Smartbody is made through Stomp library (Briggs, 2018). We use the Behavior Markup Language (BML) (Kopp et al., 2006) standard to synchronize between the multiple gestures. Each behavior can be marked according to the timing (ready, start, stroke, relax, end) and amount (amplitude) of the gesture, following the work of Kendon (Kendon, 1997).

The behavior generation process includes multiple steps to analyze, select and schedule the non-verbal gestures aligned with the speech of the agent: natural language processing, rule-based behavior selection and emotional dynamics post-processing. Figure 6.1 shows the overall architecture of our system. The system takes the desired spoken utterance of the agent as text input, along with parameters that show the expressiveness and dynamics of emotional regulation. The latter parameters consist of the characteristics of the agent, which is independent of the input text. The input text is first parsed and tagged in the natural language processing step. The tagged text is then assigned a set of gestures with rule-based gesture selection and gesture refinement processes.

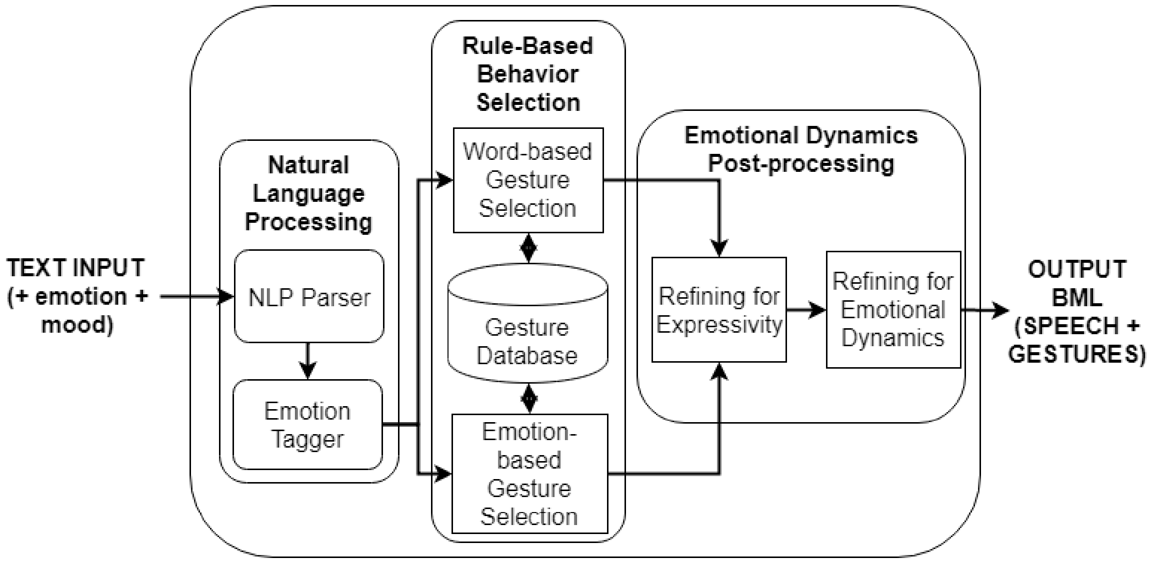

Figure 6.1: The framework of the overall model consists of the processing steps from the textual input to the output of synchronized gestures with speech.

In order to evaluate the emotional communication competency of our agent and its extent of affective expressiveness, we first conducted a series of experiments that show the emotional expression repertoire of the agent. This step will help us to correctly relate the gestures of our agent to the parameters of our system. The next section will describe this process before going over the modules of our framework.

6.5.1 Gesture Repertoire of the Agent

To determine our agent's emotional expression repertoire, we conducted initial experiments with our ECA. For this, we selected a set of natural body gestures and facial gestures that correspond to a set of emotional tags. Using the amplitude, acceleration and duration of each gesture, we are trying to understand their effect on the perceived emotion in Pleasure-Arousal-Dominance (PAD) scale (Mehrabian, 1996) as well as Ekman's basic emotion categories (Ekman, 1992). All of the calculations were done in R (R Core Team, 2018) environment, using statistics and plotting packages lme4 (D. Bates, Mächler, Bolker, & Walker, 2015) and ggplot2 (Wickham, 2016).

Each gesture is analyzed based on the temporal and positional characteristics (see Figure 6.2). Timing features include the "ready", "start", "stroke", "end" and "relax" synchronization points, similar to the annotation scheme of Kipp and colleagues (Kipp, Neff, Kipp, & Albrecht, 2007). The duration of the gesture marks the time between the start and end of the gesture, where the peak of the gesture marks the stroke. The amount of the spatial movement shows the amplitude of the gesture, where the maximum change occurs during the stroke time. The speed of the gesture can be calculated both on acceleration (from start to stroke) and deceleration (from stroke to end) of the gesture.

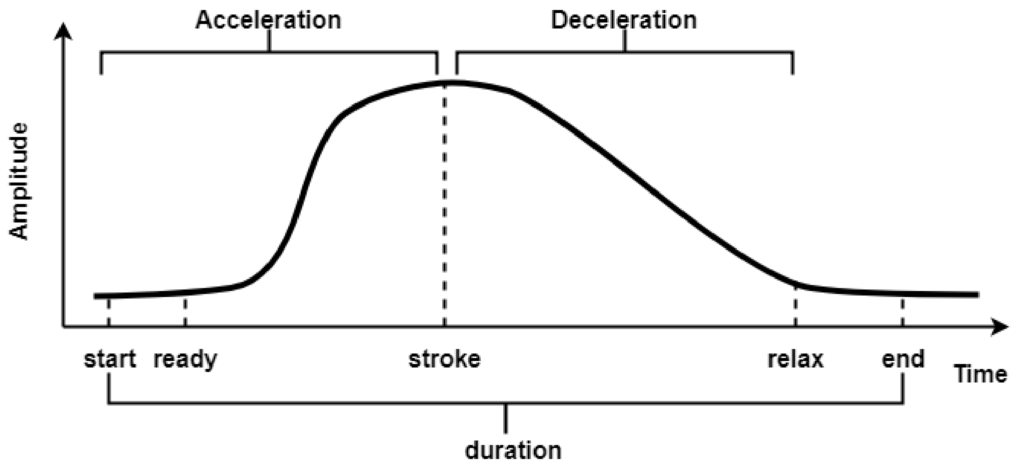

Figure 6.2: The temporal and spatial characteristics of the gestures. Maximum amplitude of the gesture is reached at stroke time.

The gestural repertoire of the agent is used to generate a lexicon for the automated gesture generation model via an evaluation process (see Figure 6.3). Short video recordings of each facial and bodily gesture of the avatar with different temporal and spatial values are collected in a database. These videos are then used to evaluate the perceived emotional value of each gesture by subjects. The facial gestures and bodily gestures are evaluated separately using a crowdsourcing platform. The results are then used to generate the model to be used in the automated gesture generation system. This process for evaluating the face and body gestures is explained in detail in the following sub-sections.

Facial Gestures

Facial gestures of the agent include AUs related to the expression of the emotions. In order to prepare initial facial gestures for our agent, we used Ekman and Friesen's (Ekman & Rosenberg, 1997; Ekman, 1972) Facial Action Coding System as a baseline for generating emotional expressions. Table 6.1 shows an overview of the selected AUs for the emotional expressions for our avatar. The expressions are refined according to the naturalness of the gesture in the lab environment before the evaluation of each gesture.

The expressive accuracy of the facial expressions of the agent evaluated for the purpose of emotion expression. For our automated gesture generation system, we examined the accuracy of the facial expressions of basic emotions that are expressed by the 3D avatar in the Smartbody system. In order to achieve this, we prepared 5 second recordings of the avatar expressing basic emotions: happy, sad, surprise, fear, anger, disgust and contempt. We used different amplitude values for each emotional expression to compare subtle differences. A total of 3 videos (1, 0.7 and 0.4 amplitude) prepared for each of the seven emotions. Some

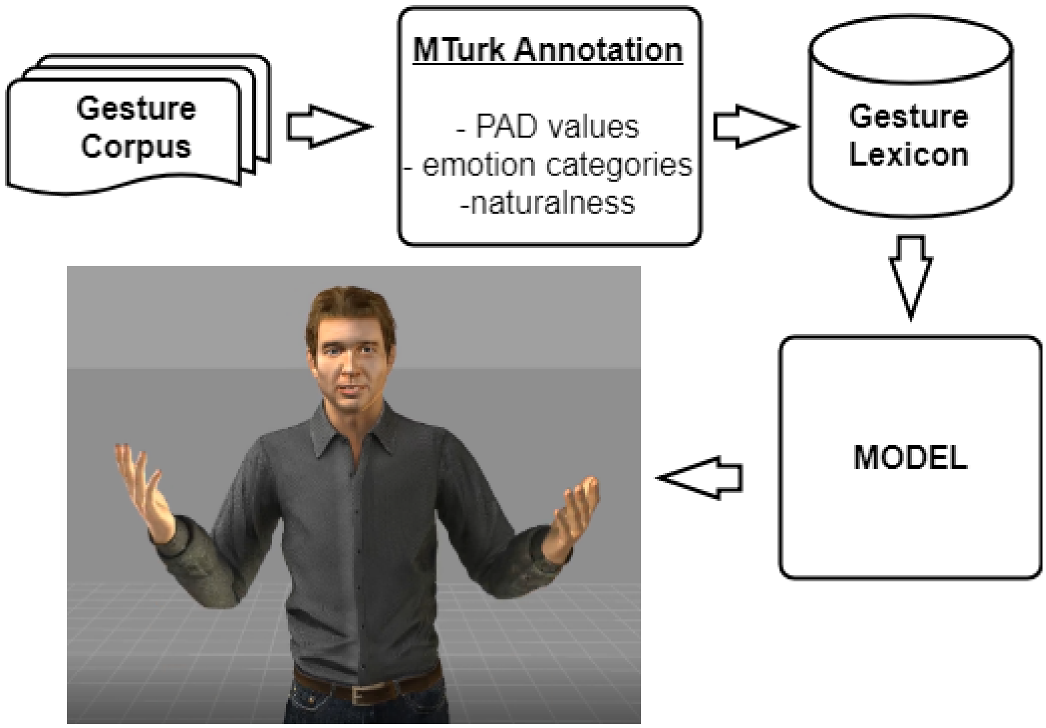

Figure 6.3: The gesture model generation process. Videos for each gesture in the repertoire of the agent is analyzed by crowd-sourcing. The data collected from this process is used to generate a gesture lexicon, which is used for generating the gesture model.

of the emotions (happy, fear, sad) had alternative expressions for testing purposes. We used Mechanical Turk crowdsourcing platform to evaluate these videos in terms of perceived emotion in Ekman categories (Ekman, 1992) and Pleasure-Arousal-Dominance dimension values (Mehrabian, 1996). In order to assess the validity of the participant responses, we used filler questions that determines whether or not the participants are paying attention to the videos.

As it can be seen on Table 6.2, anger (%52), happy (%68.6), sad (%38) and surprised (%61.5) are categorized with better accuracy. Other emotions such as contempt (%16.6), fear (%10.8) and disgust (%33.3) were mostly falsely recognized as other emotions. However, most of the misrecognized videos seem to have low amplitude. When only comparing the videos that represent high amplitude expressions, we see that the correct percentage increases: anger (%66.7), happy (%88.9), disgust (%55.5), sad(%55.6), fear (%33.3) and surprised (%55.6). It should also be noted that these videos were shown out of context, which might have contributed to the low recognition scores. For example, fear and surprise share very similar AUs which can explain the confusion between the expressions. This misrecognition issue has also been noted in various behavioral studies (Roy-Charland, Perron, Beaudry, & Eady, 2014). Similarly, contempt can be seen as an asymmetrical expression of happiness and can be misinterpreted as happy in behavioral studies (Shioiri, Someya,

Table 6.1: Action Units used for Emotion Expressions

Emotions	AUs	Alternative
ANGER	4, 5, 7, 23	2, 4, 5, 7
CONTEMPT	R12A, R14A	-
DISGUST	9, 15, 16	-
FEAR	1, 5, 26	1,2,4,5,7,20,26
HAPPY	6, 12	6, 12, 1, 2, 5
SAD	1, 4, 15	1, 4, 6
SURPRISED	1,5, 26	1, 2, 5, 26

Table 6.2: Emotion Comparisons for Expressive Accuracy

Emotions	ANGER	CONTEMPT	DISGUST	FEAR	HAPPY	SAD	SURPRISED
ANGER	**13**	4	1	1	1	3	2
CONTEMPT	3	**4**	1	0	***15***	1	0
DISGUST	3	5	**8**	0	0	7	1
FEAR	2	6	3	**4**	3	8	***11***
HAPPY	6	7	0	0	**35**	2	1
SAD	5	7	3	3	6	**18**	5
SURPRISED	2	2	0	2	2	2	**16**

Helmeste, & Tang, 1999). A comparison of the AUs used in the study can be seen in Table 6.1.

We have also evaluated the effect of amplitude on perceived PAD values of the emotional expressions. Results of ANOVA showed that there was a statistically significant effect of the amplitude of facial gestures on Pleasure (F(2, 253)= 4.074, p=0.0181) and Arousal (F(2, 243)=2.95, p=0.0542). However, no significant difference was found on the effect of Dominance (p=0.3). Figure 6.4 shows a graph of valence for each emotion, with different amplitude values. As expected, the positive emotions increased the Pleasure category, while negative emotions showed a decrease in valence when the amplitude increases. One outlier is "contempt" emotion, which was found to be misrecognized as "happy" emotion in the previous study. On the other hand, arousal values show a steady increase in all of the emotions (see Figure 6.4), with slight differences in "shocked/surprised" and "anger" emotions.

Body Gestures

The body gesture repertoire of our agent consists of a set of behaviors that can be categorized as Iconic, Deitic, Metaphoric and Beat gestures, following the categorization of McNeil (McNeill, 1992). Each gesture can include a combination of hand, body posture, shoulder, torso and head movements. The Beat gesture category includes Left-Handed, Right-Handed and Both-Handed versions, as well as variations on amplitude and speed. In order to generate

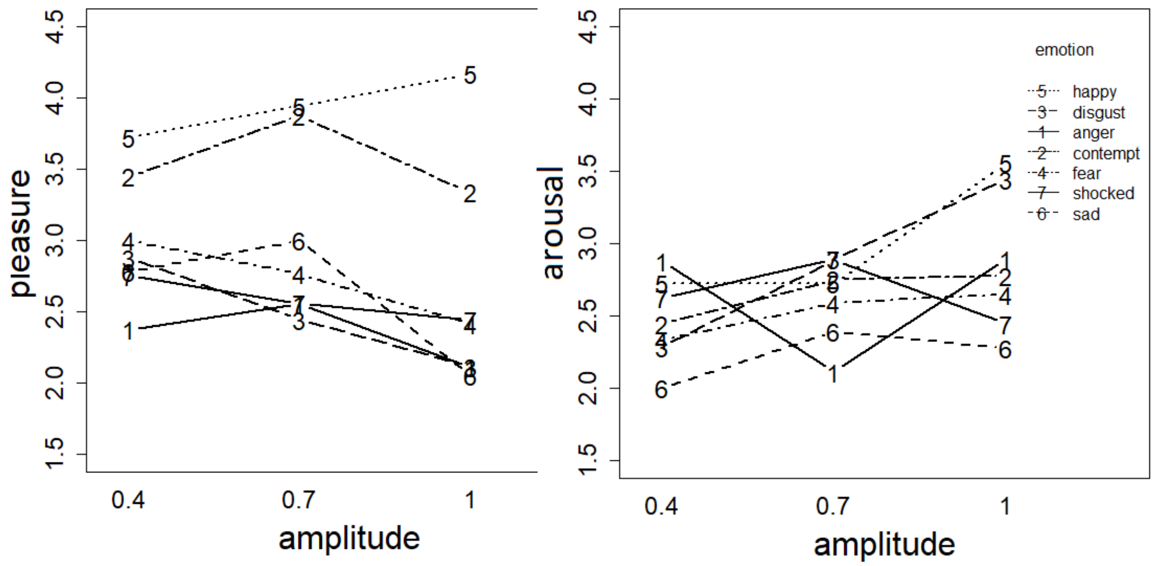

Figure 6.4: The interaction between gesture amplitude and the perceived Pleasure-Arousal values for each gesture. Both figures show anger (1), contempt (2), disgust (3), fear (4), happy (5), sad (6) and shocked/surprised (7) emotion categories.

a model for our system, we needed to compute the amplitude, duration and speed variables for every gesture phase (start, stroke, end) of every gesture.

In order to achieve this, we collected the motion files of every gesture in the agent's repertoire. Smartbody provides "skm" files for every gesture, that shows the specifications of the rotations of each joint in the agent's skeleton. These files include quaternion values for every thirty milliseconds and needs to be converted to positional values. In order to calculate this information, we first converted ".skm" and ".sk" files to ".bvh" standard. After this step, we used a modified version of PyMo motion calculator library (Omid Alemi and Ozge Nilay Yalcin, n.d.) to calculate the total position change at every phase during the motion. After collecting these values and the timing information for each gesture phase, the speed of the gesture between phases can be calculated[1].

After collecting this information, we conducted another study using crowd-sourcing to determine the PAD values perceived for each gesture. We examined the effect of duration, acceleration and amplitude of the gestures into PAD scale. We conducted a study with 1890 ratings for a total of 118 gestures.

We analyzed the pleasure, arousal and dominance values to be able to find the correlations between amplitude, duration and speed variables of each gesture. In order to achieve this, we used linear mixed models (D. Bates et al., 2015) to be able to understand the

[1]All the results from these calculations as well as the converted files will be made publicly available at https://doi.org/10.48410/zr57-m463

effects of multiple variables to each PAD dimension separately. Results showed that amplitude (beta = -0.16, SE = 0.081, $t(916) = -1.94$, $p = 0.05$), duration (beta = 0.73, SE = 0.11, $t(916) = 6.91$, $p < .001$) and speed (beta = -0.17, SE = 0.038, $t(916) = -4.54$, $p < .001$) have significant effect on predicting Pleasure. The effect of amplitude and speed was negative, where duration had a positive effect.

Similarly, amplitude (beta = 0.44, SE = 0.094, $t(916) = 4.68$, $p < .001$), duration (beta = -1.25, SE = 0.12, $t(916) = -10.21$, $p < .001$) and speed (beta = 0.22, SE = 0.044, $t(916) = 4.99$, $p < .001$) have significant effects on predicting Arousal. This time, amplitude and speed had a positive effect, where duration had a negative effect on arousal values. Moreover, there was a significant effect of duration in dominance (beta = 0.40, SE = 0.13, $t(916) = 3.06$, $p < .01$). We found no significant effect of amplitude and speed on dominance ($p > .1$ for both cases).

The results from Pleasure and Arousal were as expected. Yet, we were expecting to find a significant positive effect of amplitude and speed on dominance, following the known research on this topic.

6.5.2 Natural Language Processing

The first processing step of the system analyzes the textual input using multiple natural language processing tools. These include the parser and the emotion tagger components. The parser is used for splitting and tagging the text into sentences and words, to be able to prepare the speech BML message. It is important to tag every word in order to synchronize the gestures in relation to the timing of every word and even phoneme. This process allows triggering any gesture at the beginning, end, and any relative time from these points in seconds. After parsing is complete, every word and the whole sentence is then analyzed and tagged for its emotional value. The emotional value of the sentence is a sentiment value between positive (+1) and negative (-1) sentiment. Additionally, each word has its emotional value in terms of emotion categories, the strength of each emotional value and Pleasure-Arousal-Dominance (PAD) value.

Our system uses NLTK library (Loper & Bird, 2002; Bird, Klein, & Loper, 2009) and Stanford's CoreNLP (Manning et al., 2014) for the initial parsing and tagging. The emotion recognition system uses the SO-CAL sentiment analyzer (Taboada et al., 2011) with NRC-Canada System (Mohammad et al., 2013). The emotion recognition process is explained in more detail in previous work (Yalçın & DiPaola, 2018).

Figure 6.5 shows an example of this process for the input text "I feel amazing today". The emotion recognition is only used with categorical parameters for the example, but PAD recognition and sentence sentiment is also calculated for each input. Parsed and tagged text is then sent to the rule-based gesture selection module to initiate the gesture selection process.

```
input = "I feel amazing today."
parsed_text=[
    Mark(name=T1), 'I', Mark(name=T2),
    'feel', Mark(name=T3), 'amazing',
    Mark(name=T4), 'today', Mark(name=T5)
    ]
emo_val=[None, None, {'value': 0.788, 'emotion': 'joy'}, None]
```

Figure 6.5: An example output of the natural language processing step of an input sentence. The sentence is parsed and emotion recognition values for each word is computed.

6.5.3 Rule-Based Gesture Selection

After receiving the parsed and tagged text from the natural language processing module, word-based and emotion-based gesture selection are used to generate semantically-related gestures synchronized with the text. The word-base selection of gestures ensures the usage of initial iconic, metaphoric and deitic gestures in relation to the words and concepts present in the provided text. These rules that specify the association between non-verbal behavior to speech are stored in the knowledge base of the agent. These behaviors are related to the objects and actions, spatial relationships and relations between concepts as well as the special emphasis words the agent is programmed to react towards. For each word or phrase that has a matching gesture representation, a BML message for that gesture is generated to be triggered at the specific time of that word.

Emotion-based gesture selection relies on the tagging of affect-related components of the text, which consist of the word-based emotion values. At this step, the facial gestures related to each emotional word are included in the BML with matching timing on the word. Note that there might be several matching emotional gestures associated throughout the text, where the timing and amount of the emotional behavior need to be refined at further processing steps.

```
Gesture(name=ChrBrad@Idle01_MeLf01, stroke=speech:T1)
Gesture(name=ChrBrad@Idle01_HereBt01, stroke=speech:T4)
Face(emotion=happy, amount=0.79, stroke=speech:T3)
```

Figure 6.6: The result of the rule-based gesture selection process following the previous input example. Three gestures selected from the database based on words and emotions within the input.

An example of this process can be seen in Figure 6.6 for the text "I feel amazing here". The word-based gesture generation picked a deitic gesture to signify "here" and iconic gesture for "I". Note that the timing for the stroke of each gesture is aligned with the signified word.

Affect-based gesture generation for facial gestures uses the input from the previous module for the affective properties of the sentence. Aligning the stroke with the emotional signifier word "amazing", the happy gesture for facial expression is selected with the valence value provided for that emotion. This value is going to represent the amplitude of that emotional gesture. These values provide a guideline for the further refinement of the gestures but can also be used to generate behavior at this point.

6.5.4 Emotional Dynamics Post-Processing

The last step for the gesture generation system involves the refinement of the gestures received from the previous step in terms of the dynamic properties of the agent and the utterance. This post-processing module consists of the last stage of processing in the system that refines the gestures selected to accompany the speech according to the desired expressiveness and emotional dynamics. The input sentence, emotional values and the proposed gestures generated from previous processing steps are used as an input for this module. The output of this processing step depends on a set of parameters that defines the behavior of the agent in terms of emotional dynamics: expressivity, the degree of variability and time-dependency.

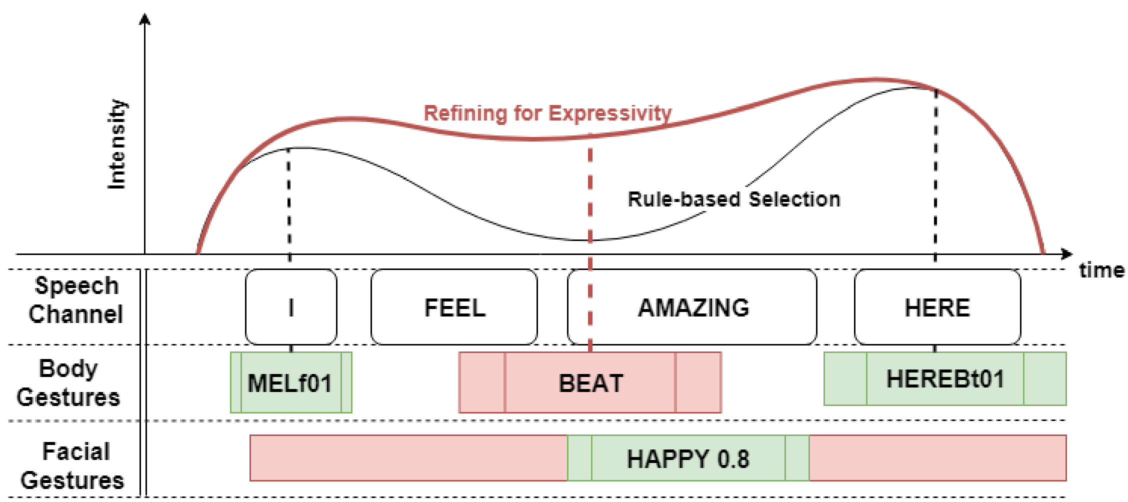

Figure 6.7: Figure shows the first two stages of the emotional dynamics post-processing module. The inputs received from the previous module is colored in green. The gestures and properties shown in red are added according to the expressivity, and emotional dynamics variables.

Expressivity determines the agent's tendency to use gestures during the utterance in terms of frequency. This allows for a consistent pattern of behavior while contributing to the sense of personality in the agent. This parameter is not subject to change during the interaction and can be re-set with every new character. According to the expressivity value of the agent, the selected gestures from the rule-based gesture generation step can be removed or increased, changing the density of gestures along the speech act.

Following the previous example on gesture generation, the input received from the rule-based generation component included bodily gestures that were signifying 50 percent of the speech text. If the agent's expressiveness parameter is between .75 and 1.0, the post-processing step would have selected an additional BEAT gesture and placed it in the least dense position in relation to other gestures. Figure 6.7 shows how the addition of this extra gesture effects the output. If the agent's expressiveness parameter would have been less than .5, then one gesture from the densest position would be picked for removal. In our example, where both gestures are equally distributed along the speech behavior, both gestures would have an equal probability of removal. This can be changed by assigning priorities to the gestures in terms of their types (iconic, deitic, metaphoric). By default, beat gestures have the least priority in our system.

The second stage of behavior post-processing includes refining for emotional dynamics. This process uses the input received from the previous steps as well as hard-coded parameters of the degree of variability and time-dependency, following the work of Kuppens and Verduyn (Kuppens & Verduyn, 2015). After the initial processing of bodily gesture expressivity, these variables are used for determining the shape of the emotional face gestures. These variables determine the duration and stroke intensity of the emotional gesture. Figure 6.7 shows with a minimal time dependency, the happy gesture that is shown in green can be expanded throughout the whole speech behavior (shown in red). This allows the gesture to be sustained for a long time unless another emotional gesture intervenes with the process. Similarly, the degree of variability determines how much the gesture will be held in its maximal intensity during the stroke. The interaction between these variables may contribute to the perception of mood in the agent.

This stage also makes use of the overall sentiment of the sentence in order to control for contrasts and multiple emotions throughout the sentence. An example of this can be seen in Figure 6.8, with a sentence that accompanies words that signify happy and fearful emotions. The baseline gesture generation for this sentence from the previous stage is shown with the green and red lines/boxes, that are centered around the emotional words. The post-processing step for this sentence will refine these gestures where the overall sentiment is skewed towards negative. This will result in a drop in the activation of happy gestures and a trajectory towards the negative expression, which is fear.

As a result of both of these processing steps, our system outputs a BML that signifies the selected gestures with the refined spatial and temporal properties. This BML output is then sent to the Smartbody behavior realizer to be executed as animations for the 3D anthropomorphic agent.

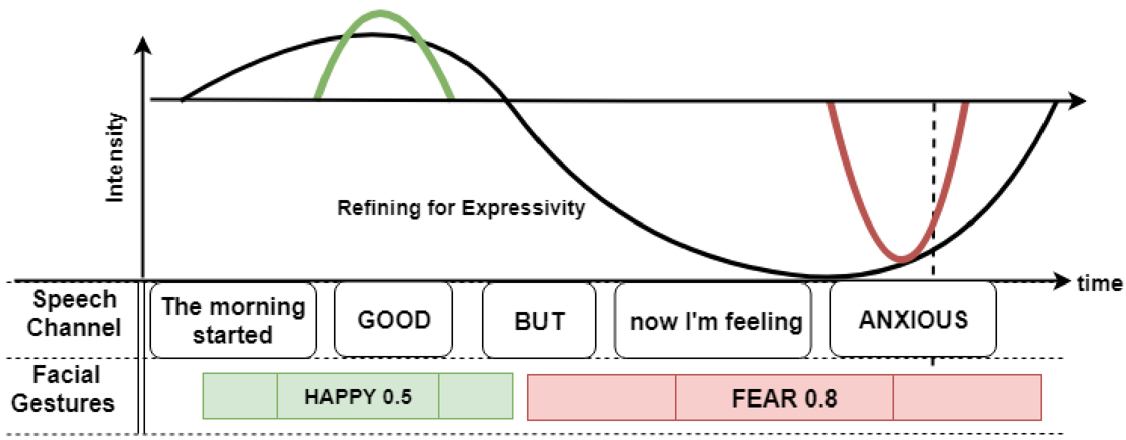

Figure 6.8: An example gesture post-processing for a sentence that includes contrasting emotions. The green and red colored lines show baseline gesture generation, where the black curved line show the result of the post-processing step.

6.6 Conclusion

Emotional expressions during speech require the synchronization of multiple behavior channels. Automated generation of emotional behavior that is both expressive and fluent is a challenge in real-time embodied conversational agents. In this work, we propose a system to find a natural balance between the semantics, intrinsic qualities of the agent and affective dynamics of the utterance. We showed a three-staged framework to process the semantic and affective qualities of the utterance, generate a baseline from this information and finally refining the gestures based on the properties of the agent and the emotional dynamics. For future work, our trajectory is to extend the speech gestures to proper listening behavior and apply this system in a real-time conversational framework.

Chapter 7

Levels of Emotional Contagion in an Embodied Conversational Agent: Implementation and Evaluation

This paper is published in the Proceedings of the 41st Annual Meeting of the Cognitive Science Society: Yalçin, Ö. N., & DiPaola, S. (2019). **Levels of Emotional Contagion in an Embodied Conversational Agent: Implementation and Evaluation**. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pp.3143-3149.

Contributions: I was the main contributor to this paper. I was responsible for conceptualization and implementation of the system and the experimental setup. I was also responsible for conducting the experiments, collecting and analyzing the collected data as well as the visualization and writing the first draft. The development of the methodology, planning and execution were done collectively with Prof. DiPaola. Prof. DiPaola was also supervising and reviewing the paper.

In this paper, I present the evaluation of low-level empathic behaviors in an embodied conversational agent. Using the implementation of the framework described in the previous chapter, I focused in detail on the low-level (affective) empathy, which includes mimicry and affective matching behaviors. These behaviors are evaluated in a conversational setting with a human participant, where the agent responds to the emotional stories told by the participants in these low-level empathic behaviors.

7.1 Abstract

This paper presents an embodied conversational agent framework as a controlled environment to test components of empathy. We implement levels of emotional contagion which includes mimicry and affective matching along with necessary communicational capabilities. We further demonstrate an examination of these foundational behaviors in isolation,

to better understand the effect of each level on the perception of empathy in a social conversational scenario with a human actor. We report three studies where the agent shows levels of emotional contagion behavior during (1) the listening act in comparison with baseline backchanneling behavior (2) additional verbal response matching simple emotional storyline (3) the verbal response to the human actor performing complex emotional behaviors. Results revealed that both mimicry and affective matching behaviors were perceived as more empathic than the baseline listening behavior, where the difference between these behaviors was only significant when the agent verbally responded to complex emotional behaviors.

7.2 Introduction

Empathy, as the capability to understand and react to the emotions of another (Iacoboni, 2011; Coplan & Goldie, 2011), is a complex behavior that arises from the interaction of these basic affective mechanisms with higher-level cognitive functions (de Waal & Preston, 2017). Emotional contagion is said to be the foundation of empathic capacity, as it includes innate and automatic synchronization of the motor and affective responses during an interaction (Hatfield, Cacioppo, & Rapson, 1994). Behaviors such as mimicry and affective matching are levels of the emotional contagion that results from the innate capability of resonating with the other during social interaction.

The literature suggests the sustained act of mimicry results in a feeling of the mimicked emotion and affective matching through muscular feedback (Hatfield et al., 1994; Hatfield, Bensman, Thornton, & Rapson, 2014), while categorizing both behaviors as emotional contagion. Others use affect matching as a highly connected but distinct phenomenon to the mimicry, pointing out the differences between the subjective quality of experience in the emotional contagion and the automatic matching of expressions in mimicry (Hess & Fischer, 2014). However, both ideas converge on the foundational role of mimicry and affect matching in empathic behavior. This notion is consistent with the Perception-Action-Model (PAM) (Preston & de Waal, 2002) and the Russian Doll model of empathy (de Waal, 2007), which integrates the neuroscience studies on mirror neurons as a baseline for the hierarchical levels of empathy mechanisms. However, it is difficult to study the levels of emotional contagion in isolation.

Research efforts often rely on behavioral experiments, neuroscientific techniques (EEG, fMRI) and pathology studies conducted to understand the effects of emotional contagion during social interactions (Hess & Fischer, 2014; Hatfield et al., 2014). As an alternative, computational empathy studies have recently gained attention in a way to simulating the empathy mechanism within the agent and examining empathic responses of the users towards the agent (Paiva et al., 2017; Yalçın & DiPaola, 2018). The perception of empathy in artificial agents is shown to increase the length of the interaction (Leite et al., 2014), user performance (Partala & Surakka, 2004), user satisfaction (Prendinger et al., 2005), and lead to more trust (Brave et al., 2005). These findings suggest that equipping interactive systems with empathic capacity would not only improve our understanding of the interaction between cognitive and affective processes in the human mind but may also help us enhance our interaction between artificial systems.

In this work, we use the simulation approach to study empathic behavior in virtual agents and try to understand the differences between the levels of emotional contagion behavior and the perception of empathy during a conversation. We examine the basic emotional contagion capabilities in an embodied conversational agent (ECA) in order to evaluate the perception of empathy during mimicry and affect matching behaviors. We present an agent framework and implementation with necessary communicational capabilities as a baseline. In the following section, we will present our implementation for an ECA that incorporates different levels of emotional contagion as a foundation for empathic capacity. Next, we will demonstrate three experiments that examine the effect of these levels on the perception of empathy during a social interaction scenario with a human actor. Our approach and results show the potential of computational empathy studies as a reliable alternative to test mechanisms for empathic behavior in isolation.

7.3 Agent Behavior

Our empathy framework is implemented in an embodied conversational agent that is capable of responding to an emotional conversation with the user using verbal and non-verbal behaviors. Our socially situated 3D virtual character system can perform a set of behavioral acts and context-specific dialogue in response to the speech and video input received from the user (see (Yalçın, in press) for a detailed explanation of the framework). Inputs are gathered using a standard webcam and a microphone. We use the Smartbody behavior realizer (Thiebaux et al., 2008), that can provide face and body gestures, gaze, and speech output for virtual characters. We use the standard Behavior Markup Language (BML) (Kopp et al., 2006) as the basis for the two way connection between the framework and the behavor realizer.

The implementation includes mimicry and affect matching behaviors as the foundational capabilities of empathy in combination with basic conversational capabilities such as backchanneling. In order to achieve this, our system incorporates a perceptual module, a behavior controller and a behavior generation module. The visual and verbal input from the user is processed through the perceptual module, reasoned within the behavior controller according to the selected empathy mechanism and prepared for a behavioral output in the behavior manager before being displayed in the ECA.

Low-level empathic behaviors, such as mimicry and affective matching require a fast response to the emotional stimuli presented by the interaction partner. The fundamental components of this first level of empathic behavior include the perception of emotion, representation of emotion and expressing emotion. This cycle is realized with Perceptual Module and Controller and Behavior Generation modules of our system.

7.3.1 Perceptual Module

The perceptual module is responsible for handling the input received from the user and creating internal representations of these inputs to be used by the controller. Currently, our system is capable of handling audio, video and textual inputs to be used in recognition systems. The audio input includes verbal signals from the user to be recognized as speech and pauses. The initiation, pauses and termination in the speech signal are used to provide information about the dialogue state as well as backchannel timing.

Emotion recognition is a sub-module within the perceptual module that is specialized for emotion recognition and fusion processes. Here, three types of modalities can be used for further processing using the first level of recognition from the perceptual module: facial emotion recognition, tone analysis and speech emotion recognition. During listening, emotion recognition is based on the facial gestures and tone analysis, which is derived from the video and speech inputs for immediate listening feedback. After the speech signal from the user ended, the complete utterance is also being processed in speech emotion recognizer for emotion detection based on the textual output of the speech recognizer. Outputs from this sub-module are used by the behavior controller depending on the dialogue state as well as the selected empathy mechanisms.

7.3.2 Behavior Controller

The behavior controller module is a central unit in the framework which provides a link between inputs and the outputs. It decides which input channel or information to be used depending on the state of the conversation, required empathy mechanisms and the behavioral capabilities of the agent. It is also responsible for providing the information necessary to the behavior manager module to prepare verbal and non-verbal behavior. The Controller acts as a decision-making component, which determines behavioral choices concerning the percepts of the agent and its internal state. Currently, the behavior controller provides a link between the perception-action mechanisms as a key component in computational empathy (de Waal, 2007). During a conversation, the agent should decide which behavioral state it is in depending on the user input: listening, thinking, speaking or waiting. According to the state of the interaction (listening, speaking, thinking and waiting) and the current emotional value (arousal, valence and emotion category), the controller assigns the proper behavior categories to the behavior generation component.

If the user is speaking, the agent should be in the listening mode. Here, the agent is expected to provide proper backchanneling to the user as well as the emotional feedback depending on the empathy mechanisms. After the speech of the user is over, the speech signal should be sent to the dialogue manager component through the controller with the assigned emotional value. The agent will be in thinking mode during the processing of this input by the dialogue manager component. The prepared output sentence will then be sent back to the controller to be sent to the behavior manager which will prepare the output behaviors including face gestures, body gestures and the verbal response to be presented in the speaking mode. After the speech behavior of the agent is done, the waiting or idle mode will be activated until the user speaks again. This cycle can be interrupted via the controller at any stage.

7.3.3 Behavior Generation

The behavior generation module is responsible for preparing the output for the virtual character depending on the emotion, dialogue state and speech information received from the behavior controller. During listening behavior, this module is relatively passive in preparing behaviors. It uses the backchanneling signal to select an appropriate head nod for the agent and a facial expression. When these behaviors are sent and consumed by the behavior realizer, the behavior generation module receives a signal back that indicates the behavior was successfully generated by Smartbody system.

7.4 Method

In this paper, we used the simulation approach to study low-level empathic behavior in virtual agents to show the differences between the levels of emotional contagion behavior in the perception of empathy. We examined the effect of mimicry and affect matching behavior on perceived empathy during conversational interaction using three studies.

7.4.1 Participants

Participants for all three experiments were recruited using Amazon's Mechanical Turk platform and were paid for their participation to the study. Because we were focusing on the emotional expressions during verbal communication, we only included participants who had English as their first language. Additionally, users that participate with mobile devices and tablets were excluded to ensure a consistency in the display quality.

A total of 84 subjects participated in the studies. 36 participants with ages ranging from 20 to 60 (M=37.6, SD=10.7) completed the first study. 19 of the participants were male and 16 of them female, while 1 participant defined themselves as 'other'. 24 subjects participated in the second study with ages ranging between 21 and 64 (M=36.17, SD=10.82). 10 of the participants were female and 13 of them male, while 1 participant defined themselves as 'non-binary'. The last study included 24 participants with ages ranging between 23 and 59 (M=37.82, SD=10.64), 12 Male and 12 Female.

7.4.2 Procedure

Studies followed the same procedure, where the participants are asked to evaluate the recorded interaction between the agent and a human (see Figure 7.1). The interaction scenario consists of a student/participant expressing an emotional story to the agent. We have chosen three stories inspired by the work of Omdahl (Omdahl, 2014), that includes three basic emotion categories: anger, joy and sadness. Other basic emotions such as fear, surprise and disgust were not considered for this study due to the involvement of facial action units that controlled mouth movements during the expression of these emotions. Furthermore, we selected the emotions that would be consistent with the facial emotions, that would not provide an advantage to the affective matching over mimicry.

All of the experiments were deployed in Amazon's Mechanical Turk environment using scripts written in Python 3.6 with psiturk and jspych libraries. Each of the studies takes about 10 minutes to complete. Participants were first shown a test video and were asked to answer two questions about the visual and verbal content of the video, to make sure they can hear and see the videos that are displayed. This was required for the workers to participate in the study.

Each participant is then displayed a short video clip of an interaction, where the agent and a student are shown in a video-conferencing scenario in different conditions (see Figure 7.1). During the interaction, the student in the video talks about an emotional story in one of three basic emotions: joy, sadness and anger. After displaying each video, the participant is asked to report what the story in the video was about, and also the main emotion of the user and the virtual agent. This is done to make sure the participants are paying attention to the video clips. The participants then evaluated the perceived empathy of the agent towards the student. The perceived empathy of the agent is evaluated by using a modified version of the Toronto empathy questionnaire (Spreng et al., 2009) which is a 16-item survey that originally is used as a self-report measure. Each item on the questionnaire are scored in a 5-item likert scale (Never = 0; Rarely = 1; Sometimes = 2; Often = 3; Always = 4), where half of the items are worded negatively. Scores are summed to derive total for the perceived empathy and can be varied between -32 to +32. Similar evaluations were suggested by Paiva and colleagues (Paiva et al., 2017), as a modification of Davis's Interpersonal Reactivity Index (Davis et al., 1980).

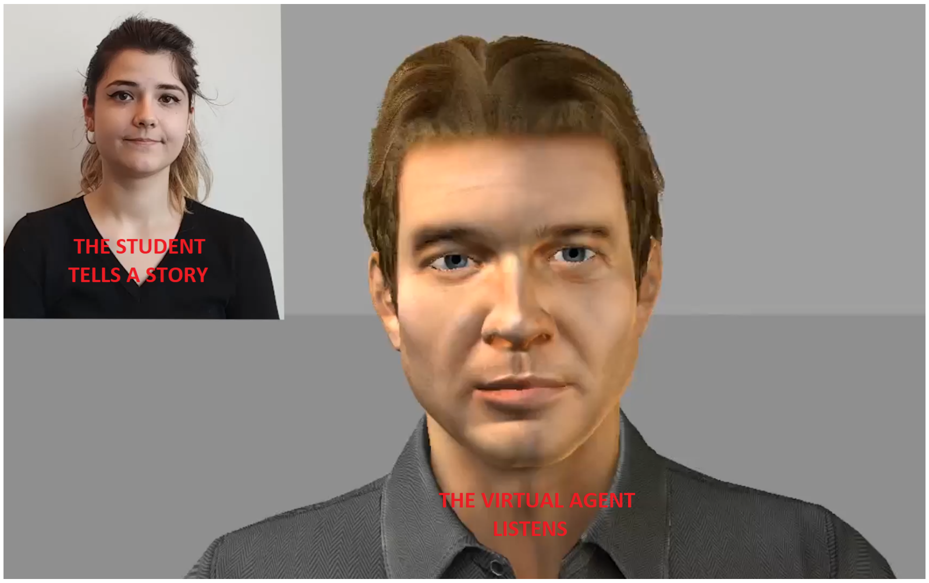

Figure 7.1: An image from the video chat between the student and the avatar. Here, the student (left) converses with the avatar.

We used repeated measures design, where each participant is shown all levels of agent behavior in emotional contagion. The type of the interaction study and the order of the conditions are counterbalanced accordingly.

7.4.3 Experiment Conditions

Experiment conditions include three distinct agent behaviors that signifies levels of emotional contagion mechanisms in the empathy framework.

The baseline behavior of the agent is the backchanneling behavior, which is activated depending on the pauses during the speech signal from the audio input component in the perceptual module. In the following subsections, we will provide a detailed examination of three different listening behaviors depending on the level of empathic behavior of the agent: backchannel only, mimicry with backchannel, affective matching with backchannel.

Backchanneling as baseline behavior

Listener behavior in humans include backchannels such as head nods, facial feedback, short vocalizations or a combination of them (Yngve, 1970). These behaviors might show information about listener agreement, acknowledgment, turn-taking or attitude (Schroder et al., 2012; Cassell, Bickmore, Campbell, & Vilhjálmsson, 2000). Backchannel feedback can occur due to change in pitch, disfluency or loudness of the speech signal, as well as shifts in speaker's posture, gaze and head movements (Maatman et al., 2005). In our current implementation we included backchanneling based on the pauses during speech, which is a form of disfluency in the speech signal (Maatman et al., 2005). Information about pauses are extracted from the perceptual module and sent to the controller, which in turn is used to trigger backchanneling as head nods. More advanced methods of adding backchannel that are compatible with the valence of the interaction partner or adding specific facial expressions such as smile, would have interfere with the empathy mechanisms that we would like to test. Therefore, we omitted these behaviors from the baseline behavior.

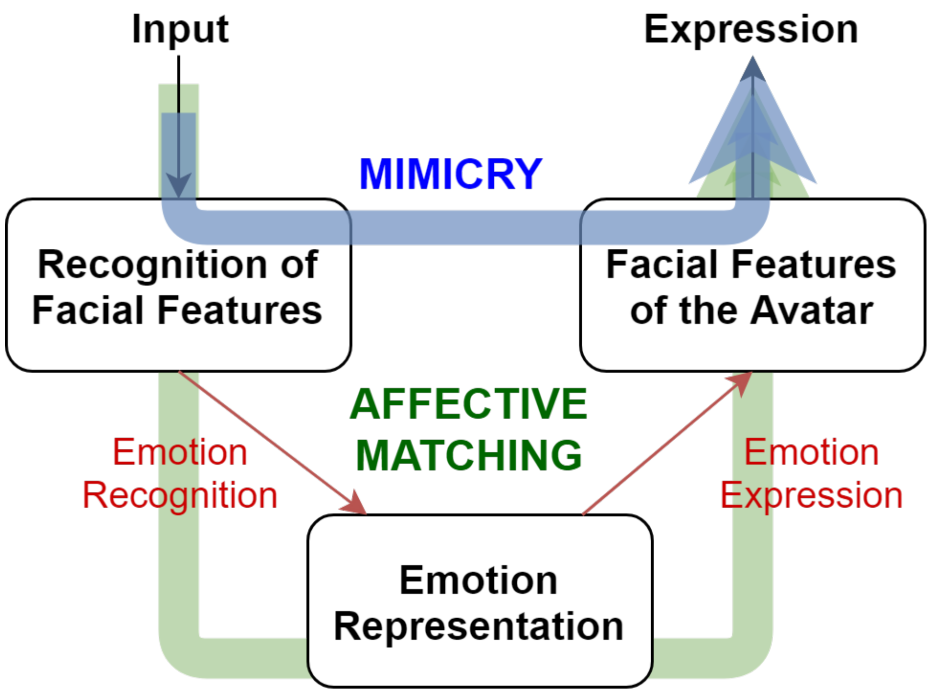

Figure 7.2: Two paths for emotional contagion. Basic emotional communication competence that results in low-level empathic capabilities of mimicry and affective matching by following distinct routes during the interaction process.

Mimicry Mechanism

Mimicry is the lowest level of empathy behavior in our empathy model. It is achieved by a direct mapping between the gestures of the user to the gestures of the agent without being assigned to any type of emotional category. Facial mimicry behavior during listening is a result of mapping the perceived facial action units (AUs) extracted from the perceptual module, to the AUs of the embodied agent in the behavior generation module. The amount, duration and speed of these AUs match the perceived values of the interaction partner without any regulations. In order to avoid mimicking of the lip movements during the speaking of the user, we removed direct mapping of AU18 (lip puckerer), AU26 (jaw drop) and AU24 (lip pressor). As a side-effect of this modification, certain emotions that requires these AUs (fear, surprise and disgust) were not properly expressed. In order to avoid bias, interactions that include these emotions were not used during the evaluation of the system for this study. However, this drawback should be noted for future studies.

After the listening cycle, the agent will sustain the mimicry behavior until it retrieves a response from the dialogue manager. The dialogue manager will then retrieve an emotionally neutral response, due to the lack of emotional representation that is needed to be acquired during the interaction.

Affective Matching

Another type of low-level or affective empathy behavior is affective matching (de Waal & Preston, 2017). It is achieved by the emotion recognition and the emotion expression cycle that is connected through emotion representation. As it can be seen in Figure 7.2, the facial features are mapped to the representation of the basic emotion categories which in turn triggers the facial expressions of the agent that represents those emotions. The amount, duration and speed of these expressions depend directly on the values from the perceived emotions. In contrast to the mimicry behavior, this allows the agent to present and regulate emotions that are better perceived by the users. Moreover, excluded emotion categories in mimicry can be used without the disturbance of the AUs that control mouth muscles as explained in the previous section.

After the listening cycle, the agent will give an emotional feedback that reflects the overall emotion of the interaction partner until it retrieves a response from the dialogue manager. In the affective matching condition, the dialogue manager is able to use the representation of the interaction partner's emotions to pick an emotional response. Without the effect of the higher level emotion regulation capabilities, the agent will pick a response that reflects the emotion of the interaction partner.

7.4.4 Study 1

In order to evaluate the perception of empathic behaviors we compared the listening behavior of the agent in backchannel, mimicry and affective matching conditions. For our study, we used within subjects design where three conditions of agent behavior are shown to the same subject for the evaluation. The conditions are baseline backchanneling behavior, mimicry with backchanneling and affective matching with backchanneling during only the listening act. We used three emotional stories told by the same person, which displays three different emotions as the main theme: joy, sadness and anger. Each video starts with a neutral remark, that is followed by the emotional story.

The experiment counterbalanced on the order of the type of interaction (backchannel, mimicry, affect matching), and the order of type of emotional story (angry, sad, happy). 36 (6x6) different conditions presented to subjects.

Evaluation

In the evaluation of the first study, Mauchly's Test of Sphericity indicated that sphericity had not been violated, X2(2) = 1.748, p = .417. A one-way repeated measures ANOVA was conducted to compare the effect of (IV) level of emotional contagion behavior on (DV) the perception of empathy in backchanneling, mimicry, and affective matching conditions. The results showed that perceived empathy is significantly effected by the type of listening feedback $F(2, 70)= 16.721$ $p < .0001$, 95%CI (see Figure 7.3). Pairwise comparisons showed backchannel feedback only (M=-5.47, SD=12.45) is perceived to have significantly lower empathy than both mimicry ($p < .001$) and affective matching ($p < .0001$). However, listening behavior with mimicry (M=5.16, SD=10.64) and affective matching (M=8.22, SD=13.72) did not have any significant difference (p = .18).

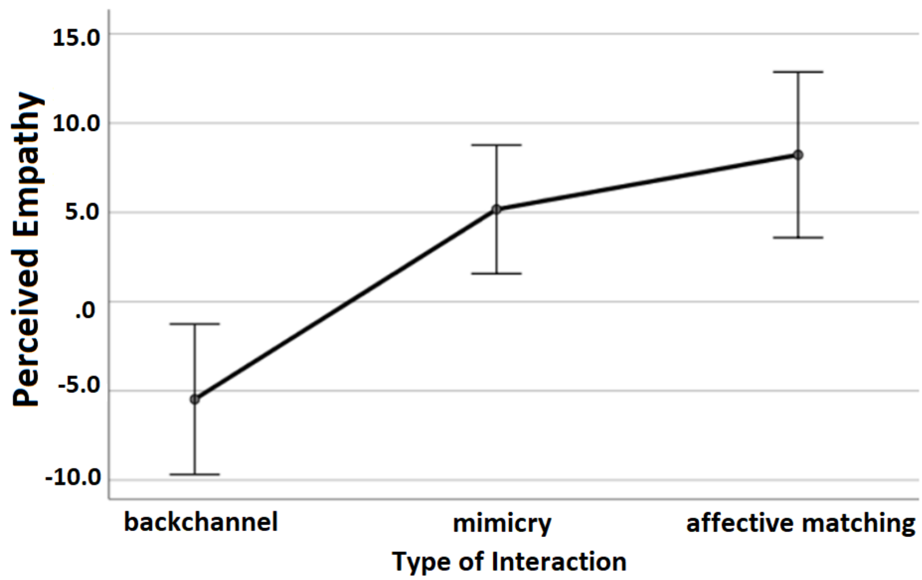

Figure 7.3: Results of our study showed significant differences in the perceived empathy levels between backchannel, mimicry and affective matching behavior (95%CI).

7.4.5 Study 2 and Study 3

Following up the first study, we further examined the effect of the verbal feedback produced by the dialogue manager in both mimicry and affect matching conditions. Our hypothesis is, due to the effect of emotional representation during the affect matching mechanism, the verbal response behavior of the avatar will be perceived as more empathic. However, this result might show difference when the interaction partner shows more complex emotions, where the context and information about the overall emotion representation is required to understand the semantics of the behavior. Therefore, we conduct two additional studies where one is focused on simple emotions and the other examines the effect of complex emotional behavior. For the following experiments, the participants were asked to evaluate the interaction stories, where the agent listens to different types of emotional stories told by the interaction partner and verbally reacts to it. As the first study showed significant differences over the baseline backchanneling behavior, the following studies did not compare the baseline behavior to emotional contagion.

In both conditions the listening behaviors of the agent will be the same as the first study, which showed no significant difference. The behavior of the agent between will differ from the first study in terms of verbal feedback during the conversational cycle. In the mimicry condition, the agent will produce an emotionally neutral feedback such as "I understand" or "I know what you mean" while sustaining the reflective facial expression of the interaction partner. In the affect matching condition, due to the additional information the dialogue manager will receive from the emotional representation of the interaction partner, the agent will produce an emotionally charged sentence. The emotional category of this sentence will be the same as the emotions of the interaction partner. For example, a happy story will trigger a happy remark such as "That sounds wonderful", an angry story will trigger a response such as "That is really frustrating", and a sad story will trigger a sad response such as "I am sorry to hear that".

The third experiment focused on more complex emotional stories, where the human actor will talk about two scenarios mentioning a dog and a plant. In the dog scenario, the actor will go through excitement, disgust, worry and happiness emotions while mentioning a story about their new pet dog. In the plant scenario, the actor will go through neutral, surprise, worry and happiness emotions while mentioning a story about their friend's plant. The listening behavior of the agent will be matching the emotions both in mimicry and affective matching conditions. Similar to the second study, mimicry condition will result in a generic verbal response from the agent while affective matching condition will give an emotionally charged feedback due to emotional representation.

The second experiment counterbalanced on the order of the type of interaction (mimicry, affect matching), and the order of the type of emotional story (angry, sad, happy). 12 (2x6) different conditions presented to subjects. The third experiment is also counterbalanced on the order of the type of interaction (mimicry, affect matching), and the order of the type of emotional story (dog, plant). 4 (2x2) different conditions presented to the subjects. Both experiments followed the same procedure as the first study.

Evaluations

In the second study, one-way repeated measures ANOVA was conducted to compare the effect of (IV) level of emotional contagion behavior on (DV) the perception of empathy in mimicry, and affective matching conditions. The results showed that perceived empathy is not significantly different between mimicry ($M=7.62$, $SD=11.66$) and affect matching ($M=9.5$, $SD=8.03$) conditions $F(1, 23) = 1.030$, $p = .321$.

Following up these results, in the third study, one-way repeated measures ANOVA was conducted to compare the effect of (IV) level of emotional contagion behavior on (DV) the perception of empathy in mimicry, and affective matching conditions during the interaction with complex emotional behavior. The results showed that perceived empathy is significantly different between mimicry (M=0.75, SD=10.45) and affect matching (M=7.21, SD=9.98) conditions $F(1, 23) = 7.731$, $p = .011$ (see Figure 7.4).

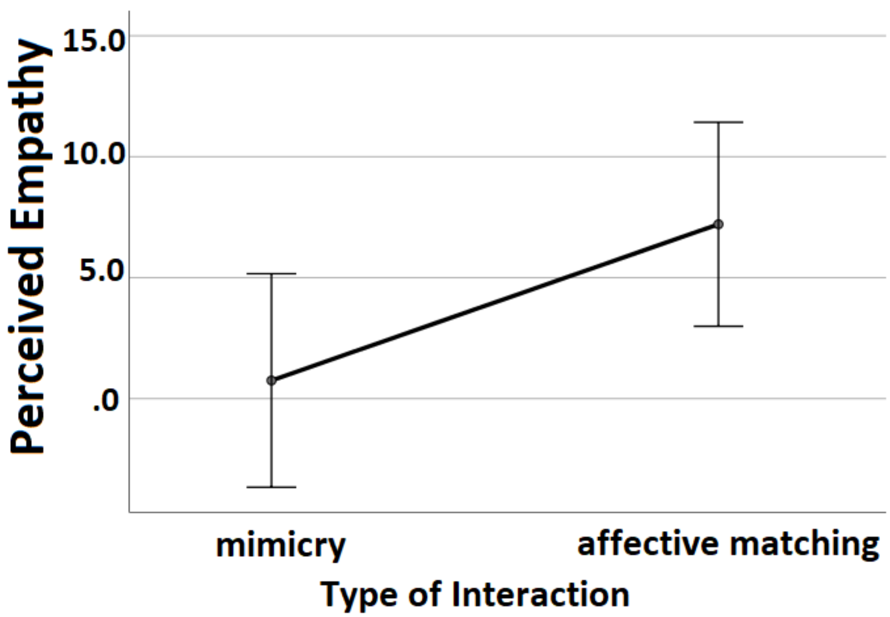

Figure 7.4: Results of the third study showed significant differences in the perceived empathy levels between mimicry and affective matching behavior in complex emotional interaction (95%CI).

7.5 Discussion

The results of the studies showed a significant difference in the perception of empathy between the baseline backchanneling behavior and the emotional contagion behavior during the listening act. As expected, the perceived empathy was significantly higher in the emotional contagion behavior with respect to the baseline behavior in the first study. However, there was no significant difference between the different levels of empathic behavior (mimicry and affective matching) in this experiment. This result is a direct consequence of the similarity in the expressions of these two conditions during listening.

Even though mimicry and affect matching behaviors have important differences in terms of processing of input information, the real-time expressions of these behaviors during listening behavior show dramatic similarities. During the listening act, mimicry captures the facial expressions of the interaction partner and reflects them using the same facial muscles.

In affective matching behavior, instead of copying the facial muscles, the system copies the emotions perceived from these facial expressions. As the emotions are expressed as a result of the facial muscles, these two behaviors are expected to show very similar expressions.

One advantage of affective matching that it allows the expression of emotions that are more suitable to the virtual agent, while any emotion will be expressed in terms of the virtual agent's repertoire instead of the expressions of the conversation partner. Moreover, affective matching allows processing of other input channels to conclude the emotion of the interaction partner, such as voice stress, the context of the speech and body expressions. However, the first and second studies only included simple emotions, and therefore such an effect was not present.

Another distinction between mimicry and affect matching conditions is present during the verbal response after the listening act is completed. This response is created by examining the overall emotion of the story told by the interaction partner. As mimicry behavior does not provide the representation of the emotions of the interaction partner, the virtual agent cannot generate a response that is aligned to that emotion. In contrast, the verbal response for the affective matching behavior can be generated from the emotion representation (see Figure 7.2 for a comparison of these two strategies). Study 2 and 3 are designed to show this distinction.

Interestingly, Study 2 did not show a significant difference between the mimicry and affect matching behaviors for simple emotional stories, where we see a significant difference in Study 3. In these studies, there are two main differences between mimicry and affective matching conditions: the content of the verbal response, and the facial emotions shown during the verbal response. In mimicry condition, the verbal response is generic where the affect matching condition generates an emotionally appropriate response. The facial expressions in mimicry response are sustained regardless of the overall emotions, where the affect matching condition generates facial expression based on the overall emotional representation for the whole story. The difference between the two studies was the emotional complexity of the overall story told by the interaction partner.

We argue that the mimicry response for the simple emotional stories in Study 2, did not show a significant difference on the perception of empathy due to the match between the overall emotion of the story and the sustained facial expression. Where in Study 3 the sustained emotion of the mimicry response was contrasting the overall emotions of the story, due to the complexity of the emotions presented by the interaction partner. We further examined the comments provided by the participants on how they perceived the behavior of the agent in response to the story told by the interaction partner. The comments of the participants in Study 2 showed that the mimicry condition is seen as “understanding" and “sympathy", where the affective matching behavior is seen as “concerned" and “empathy". In contrast, in Study 3, participant comments on the behavior of the virtual agent included descriptions such as “confused" and “indifferent", where the affect matching response was seen more as “attentive", “understanding" and “empathy". However, this distinction should be examined more systematically before reaching to a conclusion.

Overall, these results show that low-level emotional contagion behaviors of the agent during conversational interaction lead to an increased perception of empathy. Additionally, the results show that higher levels of emotional contagion behavior are perceived as more empathic behavior when the interaction includes more complex emotional behaviors. The proposed framework shows promise in providing a foundation to examine the perception of higher levels of empathic behavior during an interaction.

7.6 Conclusion and Future Work

Artificial systems provide means to test the empathy theories while allowing the manipulation of parameters in a controlled and isolated way. In this work, we proposed an embodied conversational agent framework to test empathy components and demonstrated three studies that evaluate the foundational empathy mechanisms along with basic communication behaviors. We found that during listening, mimicry and affective matching behaviors are perceived significantly more empathetic compared to backchannel behavior. We also found that the difference between the two levels of affective contagion only significant while the interaction involves complex emotional behaviors, where the context of the interaction is crucial for producing matching behavior. Our framework and the results of our initial study shows promising results that allows for easy integration and testing of higher level components of empathy. The suggested framework, study and evaluation methods shows the potential as a reliable alternative to test mechanisms for empathic behavior in isolation.

Our contributions were to provide a framework, implement the baseline behavior for real-time interaction with a highly realistic conversational avatar, and provide the first study for testing the theoretical assumptions. We hope this baseline for is useful the emerging community of researchers that study empathy in artificial agents and that it can be expanded through this framework and evaluation methods.

Chapter 8

M-Path: A Conversational System for the Empathic Virtual Agent

This paper is accepted in the Proceedings of the Biologically Inspired Cognitive Architectures (BICA) Conference: Yalçin, Ö. N., & DiPaola, S. (2020). **M-Path: A Conversational System for the Empathic Virtual Agent**. In Samsonovich, A.V. (Ed.). *Biologically Inspired Cognitive Architectures 2019. Advances in Intelligent Systems and Computing*, Volume 948, pp. 597-607. Cham, Switzerland: Springer.

Contributions: I am the main contributor to this paper. I was responsible for conceptualization and implementation of the system and the experimental setup. I was also responsible for conducting the experiments, collecting and analyzing the collected data as well as the visualization and writing the first draft. The development of the methodology, planning and execution were done collectively with Prof. DiPaola. The ideas for the topic laddering and global feedback mechanisms in the dialogue management system were initially conceptualized by Prof. DiPaola. He was also supervising and reviewing the paper.

In this paper, we present the details of the empathic conversation engine within our embodied conversational agent framework. The conversational system includes a natural language understanding component, natural language generation component and a dialogue manager that is aimed to provide a natural dialogue flow with a special emphasis to emotions. The system is intended to achieve goal-directed conversation with global goal-tracking system and local topic management system. We show the potential of our system in an example implementation as a survey conductor agent in a Psychological Counseling Service scenario. We also evaluate our system in a preliminary study with 16 users in real-time interaction. The evaluation techniques were used based on the works described in Chapter 4.

8.1 Abstract

M-Path is an embodied conversational agent developed to achieve natural interaction using empathic behaviors. This paper is aimed to describe the details of the conversational management system within the M-Path framework that manages dialogue interaction with an emotional awareness. Our conversational system is equipped with a goal-directed narrative structure that adapts to the emotional reactions of the user using empathy mechanisms. We further show the implementation and a preliminary evaluation of our system in a consultation scenario, where our agent uses text-based dialogue interaction to conduct surveys.

8.2 Introduction

Conversation forms the basis for many of our social interactions. In recent years, artificial conversational systems have become more ubiquitous and have revolutionized the nature of human-computer interaction. Natural language based assistants are becoming increasingly popular in our daily lives to accomplish goal-driven tasks and act as social companions. Emotions often provide a feedback mechanism during conversational interactions, which makes recognizing and responding to the emotions an important part of social interactions.

Empathy, as the ability to understand and respond to the emotions of others (Paiva et al., 2017), can be used as a guide to interaction. Recent examples of conversational agents in clinical psychology (Provoost, Lau, Ruwaard, & Riper, 2017) as healthcare assistants (DeVault et al., 2014) and counsellors (Lisetti et al., 2013) showed that being emotionally-aware could enhance the interaction by increasing the perceived usefulness, trust, and naturalness of the agent. These findings suggests, showing empathy during conversational exchanges can help to build and strengthen relationships while facilitating natural and believable interaction.

In this work, we aim to use empathic conversation strategies that use emotions as a feedback mechanism that helps inform the dialogue management. We describe the design and implementation of the dialogue framework for an empathic conversational virtual agent, M-Path. Our system is aimed to use system-initiated and user-initiated conversational strategies to guide the goal-oriented conversation while generating appropriate empathic responses. We further show a proof-of-concept implementation of our system in a psychological consultation scenario, where the goal of the agent is to successfully collect the required information and provide appropriate empathic responses. We conduct a preliminary evaluation that shows that our system is capable of providing empathic behavior.

8.3 M-Path: The empathic conversational agent

Our empathic conversational agent, M-Path, is aimed to create a real-time, goal-driven and empathic interaction with the user. M-Path is capable of initiating and sustaining socio-emotional interactions with the conversation partner by using different levels of empathic

Table 8.1: Example initiation of a dialogue

SPEAKER	UTTERANCE	RESPONSE TYPE
M-PATH	Hello, I don't think we met before. My name is Alex, what is your name?	Context-based (Global Greeting)
User	It's Anthony	
M-PATH	It is very nice meeting you. Are you a student here at SFU?	Context-based (Local information gathering)
User	nice to meet you too, I am indeed	
M-PATH	That's wonderful.	Intent-based (Approval)

behavior. Our embodied agent is designed to produce synchronized verbal and non-verbal behaviors to capture the richness of natural conversational interaction.

The framework for our embodied conversational agent includes a perceptual module that processes the inputs of the system via its sensors, a decision making module, and a behavior generation module that prepares and outputs the synchronized behaviors of the agent. A detailed description of this framework can be found in previous work (Yalçın, in press).

This paper focuses on the conversation engine, which is part of the decision making module of M-Path. Within the system, the conversation engine is responsible for initiating and maintaining a meaningful and goal-driven conversation with the interaction partner. This engine works closely with the Empathy Mechanisms module, that makes decisions on the empathic behavior of the agent during the conversation. In the following sections, we will examine in detail how an empathic dialogue can be simulated in a goal-driven environment and provide an implementation scenario to show proof-of-concept implementation of such a system.

8.4 Empathic Conversation Engine

The central component to achieve an empathic conversational behavior in M-Path is the conversation engine. The empathic conversation engine is designed to achieve the goal of the conversation while adapting to the emotional reactions of the user using empathy mechanisms. It consists of three major components: natural language understanding (NLU), natural language generation (NLG) and the Dialogue Manager (DM).

The Natural Language Understanding (NLU) component handles the extraction of the relevant information from the users' linguistic input and can be as simple as keyword detection to natural language modeling vector representation systems. The NLU component of our system is responsible for parsing, tagging and categorizing the linguistic input for extracting context-related information (e.g., topic, intent) and user-related information (e.g., emotions, personality, background information). This information is then sent to the dialogue manager to decide on the response of the agent.

The Natural Language Generation (NLG) component handles the realization of the intended response, depending on the decision made by the dialogue manager. Depending on

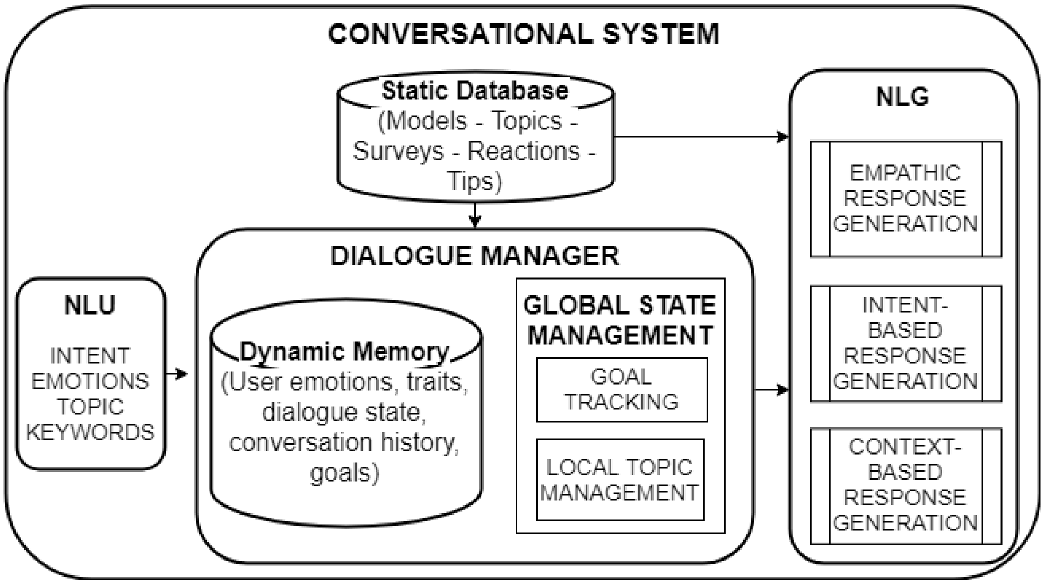

Figure 8.1: The outline of the Conversation Engine of M-Path. Three main component of the system are natural language understanding (NLU), natural language generation (NLG) and the Dialogue Manager (DM).

this decision, the response might be empathic, context-based or intent-based (Table 8.1). Intent-based responses include acknowledgment replies to the decisions or actions taken by the user. For example, if the user approves to participate in a survey, the agent will confirm that decision by “Okay" or “Thank you". Context-based responses are generated depending on the state of the conversation, both globally (goal) and locally (topic). Global state is used to determine the goal of the conversation. For example, if the user is asking a question, the goal of the conversation will be to answer that question momentarily before moving on with the main goal of the agent. The local state determines the specific topic of the utterance of the user. This is used to give topic-based answers and keep track of the hierarchical structure of the topics. For example, if the user asks “What else?", while being in the topic of “AI", the agent will continue giving examples on the topic AI.

An empathic response of the agent consists of an emotional reply to the emotions of the user, as well as a response related to the coping of the emotions. For example, if the user talks about how bad their day was, the emotional part of the empathic response could be “I am sorry to hear that" or “That is too bad". The coping response could include tips or recommendations that are less emotion-oriented and more context/topic specific. Similar to the context-based responses, these empathic responses can be local reactions to the single utterances of the user, as well as reactions to the global state of the user gathered from multiple turns of dialogue. For example, after receiving multiple negative sentiments, the agent might decide to comment on the mood of the user before saying farewell. An example for both these responses can be seen in Table 8.2 that shows a real interaction of the system with the user.

Table 8.2: Empathic reactions during conversation

SPEAKER	UTTERANCE	RESPONSE TYPE
M-PATH	How often do you feel little interest or pleasure in doing things?	Context-based
User	Very often	
M-PATH	(1)I'm sorry to hear that (2) Maybe you should try to build up a routine. You can pick up a new habit with your friends (3) How often have you been feeling down, depressed, irritable, or hopeless lately?	(1) Emotional Reaction (2) Coping Reaction (3) Context-based
User	My problem is i am not used to wake up early	
M-PATH	I understand.	Topic Fallback
User	I feel down most of the time because I do not have a friend circle in campus	
M-PATH	(1) I'm so sorry you feel like that. (2) Try not to withdraw from your social life, your family and friends. They can improve your mood a lot	(1) Emotional Reaction (2) Coping Reaction

These set of example responses are currently hand-crafted and held in a database that stores sample utterances based on context, intent and emotional reactions. Most of the contents of the static database are authored to be representative of how users might interact with the system as well as the responses added after the evaluation processes in order to be more responsive to the actual user utterances. Moreover, the Dialogue Manager (DM) also holds models that are trained based on these example utterance-response pairs. We use two main models in the DM component, where the first is TF-IDF based similarity algorithms that are used to basic Q/A. We used the sklearn library (Pedregosa et al., 2011) in Python for the TF-IDF vectorizer model, while using NLTK Lemmatizer (Loper & Bird, 2002). The second model is generated by fine-tuning the pre-trained language representation model BERT (Devlin, Chang, Lee, & Toutanova, 2018) for the classification of user responses based on sentiment and survey answers. The sequence and properties of each response are decided by using these models within the DM component.

The main role of the Dialogue Manager (DM) is to decide what action the agent should take at each turn in the dialogue. The DM is responsible for generating proper verbal responses according to the utterance of the interaction partner, as well as the empathic reactions. The dialogue manager operates on two different scales: local and global. The local structure handles the immediate responses in isolation while dealing with information gathered from individual dialogue turns. The global structure keeps track of the overarching goal of the system and operates on the entire interaction history using the local information.

Depending on the overarching goal of the dialogue and the local information, the dialogue manager decides on the proper actions to be taken. It keeps track of the context-related information (e.g., topic, intent) and user-related information (e.g., emotions, personality, background information) both locally and globally. This information is gathered from the NLU component and stored in the Dynamic Memory within the DM.

8.4.1 Global State Management

The global state tracking sub-module keeps track of the overarching goals of the agent in order to decide on which actions to take in a given dialogue turn. Every dialogue is embedded in a generalized structure of greetings/introduction, main topic flow and conclusion/closings. For example, the initiation of every conversation starts with a simple greeting, introduction and basic questions that can be used as a transition before going through the main topic. Similarly, the concluding remarks may include an overall statement about the dialogue and a farewell. The concluding statement can include an empathic response to the overall mood of the user during the conversation. An example excerpt for the introduction and conclusion processes can be seen in Table 8.1.

Goal Tracking The goal tracking sub-module operates while tracking the current state of the goal or multiple goals, which in turn determines how the main topic flow operates. Goals usually include gathering certain information or going through certain topics or questions within the narrative flow. These usually trigger a system-initiated conversation, where the local topic manager allows for user-initiated conversation to allow a mixed-initiative flow. For example, if the agent has the goal of gathering demographic information, the goal tracking module makes sure the conversation flow always directs back into asking demographics questions until all the information is gathered. After this goal is achieved, the agent can move to other goals or concludes the conversation.

Local Topic Manager The local topic manager determines the flow of conversation between goal-directed conversations. Many topics make for a narrative and can be added and reused in many ways. Topic modules are a container of a narrative topic and its parts and can process through to other topics. The standard topic modules consist of a cascade of Q&A nodes, that taken as a whole create a small conversation or narrative topic. Topics can be on universal issues like “mental health” or specific to the user topics like “anxiety", “depression" and can be reused in many ways (see Table 8.3).

8.5 Example Implementation

Such an empathic conversational agent might be useful in a variety of applications, where understanding the emotions of the user and responding emphatically is important. In order to evaluate the conversational capabilities of our agent, we picked a “counselling service intake process" as an interaction scenario. For this specific use-case, we had to determine the interaction context and develop content accordingly.

Table 8.3: Local Topic Management

SPEAKER	UTTERANCE	RESPONSE TYPE
M-PATH	Would you be okay if I ask some questions about you?	Context-based Goal related Question
User	Are you a bot?	
M-PATH	Yes, are you?	Topic-Based Local QA Answer
User	No	
M-PATH	Okay	Topic Fallback
User	Sure you can ask me questions	
M-PATH	Sounds good, I will ask about generally how you feel.	Continue Context-based

To initiate a counselling service, individuals go through a screening process to gather basic information about them. This usually includes collecting demographic data, as well as a simple assessment in the form of a survey. Based on the department of the intake, these surveys vary in terms of content and questions. For the proof-of-concept scenario, we used an intake scenario for the Psychological Counselling Service (PCS) for the students of the Simon Fraser University (SFU). The screening process for the counselling office includes demographic questions and a short survey to assess the depression and anxiety of students. Demographic questions contain basic identification information such as name, age, gender, and whether or not the user is a student at SFU. The survey that follows the demographic questions is aimed to determine the severity of the issues the student may have. The common method of delivery for this intake process is a self-administered pen-and-paper form where each survey question is answered by selecting an option out of a standard likert-scale.

For the proof-of-concept implementation of our system, we picked this screening process as the main scenario of the interaction. We developed additional content for the purposes of this interaction, including a general classification system for likert-scale surveys and inclusion of the standard questionnaire as dialogue content. The goal of our agent is set to initiate the conversation, collect demographic information, conduct the intake survey and provide proper suggestions according to the survey results.

8.5.1 Development of the Material

As the material for the screening process, we used the nine items in the California Patient Health Questionnaire (PHQ) to screen for the severity of depression (Kroenke & Spitzer, 2002). Items include statements on the symptoms of depression, and patients are asked about the frequency of these symptoms they experienced over the last two weeks. The answers for the questionnaire are picked from the 4-item likert scale scored between 0 to 3, where 0 represents “not at all" and 3 indicating “nearly every day". The total score for these items indicates the severity of the symptoms, which are scored between 0-27. According to the final score of the questionnaire, there are multiple interventions that can be suggested to the patient.

To integrate this survey in our system, we used each survey question as a topic within the local topic space where the goal is to make sure all these questions are answered. Additionally, we created specific empathic responses that would serve as coping mechanisms depending on the answers of each question as well as the overall score of the whole survey.

The main goals of the agent are to gather information about the demographics of the user and to conduct a survey.Both of these goals are embedded within the global state tracking system that was explained earlier. The global state tracking system holds information about which state/goal in the dialogue the agent is in by constantly updating the state of the dialogue in the dynamic memory. This is to make sure each utterance of the agent is being evaluated according to the current intent.

For the specific implementation of the screening process, our agent can use multiple dialogue strategies in order to make sure that the goal is reached. Each of these strategies are used to make sure the agent successfully directs the conversation to reach its goals. There are two main goals of the system for the counselling scenario: gather demographics and conduct the PHQ survey. The overarching behavior of the agent within this goal-directed scenario is to act empathically towards user responses.

8.6 Preliminary Evaluation

Although the standard method of submission for these tests are the pen-and-paper survey methods, a direct comparison with this method would not be plausible due to a number of variables that is needed to be controlled. Therefore, we evaluated the empathic screening agent to its non-empathic counterpart in a text-based interaction environment. The main difference between these two agents is their responsiveness to the emotional utterances of the user. Our hypothesis is the empathic version of the conversational agent would be perceived as more empathic, which would in turn have a positive effect on the attitude towards the interaction.

8.6.1 Method

Participants A total of 16 users (10 Female, 6 Male) completed the study that were between the ages 20 and 39 (M=26.65, SD=7.74). Because we were focusing on the screening process for the student consultation service at Simon Fraser University (SFU), we chose undergraduate and graduate students in SFU. Participation in this study was voluntary and was based on open invitations to a large group of students at SFU via online communication.

Materials We used an empathic and a non-empathic version of the same conversational system to be able to evaluate the empathic properties of the system. Both versions had the same goal of gathering demographic data as well as finishing the survey for the screening process. The empathic version, as described in earlier sections, was providing emotional and coping responses based on the utterances of the user. Additionally, an empathic response was given at the end of the survey based on the overall score of the survey. On the other hand, the non-empathic version was only giving acknowledgment responses to the user utterances, and a generic closing statement after the survey is concluded. This was done to make sure only the quality of the responses are different, while the quantity of the system responses are the same.

Each user evaluated the conversational agents based on their perceived empathy. The perceived empathy of the agent is evaluated by using a modified version of the Toronto empathy questionnaire (Spreng et al., 2009), which is a 16-item survey that originally is used as a self-report measure. Each item on the questionnaire are scored in a 5-item likert scale (Never = 0; Rarely = 1; Sometimes = 2; Often = 3; Always = 4), where half of the items are worded negatively. Scores are summed to derive total for the perceived empathy and can be varied between -32 to +32.

In addition to the perceived empathy measures, we also evaluated the user's attitude towards interaction while focusing on the use-case as an alternative screening process. We used items from technology acceptance (Heerink, Krose, Evers, & Wielinga, 2009) and Godspeed (Bartneck et al., 2009) questionnaires, which includes statements about usability, believability, and human-likeness of the agent. We included items that focus on the preference towards the screening process and compares the agent-based interaction to the classic paper-based method ("I prefer the interaction to a paper-based survey") and human-initiated method ("I prefer the interaction to a survey conducted by a human"). We also included three items to understand the trust felt towards the agent, which was used in similar studies (Lisetti et al., 2013). We used a total of three statements to evaluate the trust felt towards the agent, based on "trusting the advice agent gave", "feeling better interaction privacy" and "trusting to disclose information". The total score for trust was derived from averaging these values. A 5-point Likert scale that shows agreement with the statements with items between "Strongly Disagree" to "Strongly Agree". The high scores mean more agreement with the statements where the low scores show disagreement, where the lowest score is 0 and the highest is 4 per item.

We implemented the user interface of the dialogue agents in the Slack messaging environment, where each user was using a chat channel in order to interact with the agents by using text. For the display names for the agents, we used gender-neutral names: Alex and Joe. These names were counterbalanced between the conditions as well as the interaction order and the types of surveys. This ensured there was an equal amount of participants interacting with each possible combination of agent type, order, survey type and agent names.

Procedure We used within-subject methods, where each user is interacting with both the empathic and non-empathic versions of the conversational agent. Participants used the

Slack messaging environment in standard computers in order to interact with the agents using text messaging. Participants were briefed about the context of the interaction and the procedure before the experiment. Each interaction started with an informed consent procedure.

According to the counterbalancing, each participant first interacted with one of the conversational agent (empathic or non-empathic) and took the evaluation survey about the agent after the interaction is done. After that survey, the participant went through the same process with the other conversational agent and took the evaluation survey on the second interaction. Participants had to greet the conversational agent to be able to start the conversation. Participants were assigned across conditions, while being counterbalanced in terms of the order of conditions as well as the type of survey each condition is conducting. Each subject took about 30 minutes to complete the experiment.

8.6.2 Results

From 16 users, only one encountered an unsuccessful interaction for both of the agents, where the goal of conducting the survey was not reached. None of the user responses were excluded from the final analysis of the results. All analysis and plotting are done using linear mixed models on R (R Core Team, 2018) with lme4 (D. Bates et al., 2015) package.

We performed a linear mixed effects analysis of the relationship between the perception of empathy and system type (empathic vs. non-empathic). As fixed effects, we entered the subjects into the model. Results show that perceived empathy is significantly higher in the empathic agent, relative to the non-empathic agent condition ($p = .02$).

We also examined the attitude towards the interaction. Results showed that the system type condition (empathic vs. non-empathic) significantly effects the perceived usefulness of the agent ($p = .05$). The empathic agent is found more human-like ($p < .01$) and preferred more to a human agent ($p < .01$), than the non-empathic agent. The preference of the agent over the pen-and-paper based screening process was not significantly different ($p = .2$), but high in both cases. Moreover, the results showed the system type does not have an effect on trust towards the system ($p = .41$). Table 8.4 shows details for the results.

8.7 Discussion

Results showed that the empathic dialogue capabilities that we introduced for the conversational agent resulted in an increase in the perception of empathy during the interaction in the screening process. The empathic capabilities also increase the believability and human-likeness of the conversational agent, as well as its perceived usefulness. We also see that users prefer the empathic agent more than the non-empathic counterpart in terms of its use in respect to a screening process with a human. However, we see that users would still prefer talking to a human, rather than interacting with the agent. We also saw that, counter

Table 8.4: Results of the Evaluation

Variable	Empathic agent M	Empathic agent SD	Non-empathic agent M	Non-empathic agent SD	F(1,15)	p
Empathy	3.38	8.18	-1.12	7.80	6.43	.02*
Usefulness	3.06	1.00	2.56	0.96	4.29	.05*
Human-like	2.56	0.63	1.81	0.98	10.38	<.01**
Believable	2.88	1.02	2.38	0.96	5	.04*
Prefered to human	2.06	1.24	1.69	1.40	8.99	<.01**
Prefered to paper	2.88	1.36	2.62	0.96	1.36	.26
Trust	1.81	1.02	1.64	0.95	0.71	.41

to previous studies on empathic agents, that the empathic capabilities did not increase the perception of trust.

Further examination of the scripts created from the interaction data revealed that the interactions with the agents were not homogeneous in terms of the emotions that the participants were showing. We observed that when the participants showed more negative emotions and scored lower in the surveys, they rated the behavior of the empathic agent more positively. However, we did not control for this behavior and this phenomenon needs to be examined further.

8.8 Conclusion and Future Work

In this work, we proposed and implemented a dialogue system to equip empathic behaviors in a conversational agent. We evaluated the empathic capabilities of the agent in a proof of concept use-case, the screening process in a consultation scenario. We compared the conversational agent with and without the empathic behaviors to be able to capture the effect of our system. The results suggest that the inclusion of emotional and coping responses as empathic behavior in a conversational agent leads to increase in the perception of empathy, usefulness as well as human-likeness and believability of the agent. Even though the implementation only included the PHQ survey, any type of survey can be used with a minimal amount of development process. However, in sensitive circumstances, such as the depression screening process, trusting the agent seems to pose a challenge that needs to be addressed. This system is intended to be used in an embodied conversational agent, in real-time multi-modal interaction. For future work, we intend to integrate this system into our embodied agent framework and further compare the perception of the agent during face-to-face interaction.

Chapter 9

Conclusions

9.1 Summary and Discussion

Modeling empathic capacity in interactive agents is an interdisciplinary research area that aims to enhance human-computer interaction as well as increase our understanding of empathic behavior in humans. The novelty of the field brings about a number of challenges due to the lack of standards and methods on achieving computational empathy in a variety of applications. Firstly, there is no direct translation between the well-established models of empathy from psychology and neuroscience research to computational systems and algorithms. Next, the complex nature of empathic behavior requires a multitude of systems to work together to achieve natural, real time and expressive empathic interaction. Moreover, there is a lack of validated evaluation methods that are suitable for measuring empathy in interactive agents.

Following the main research problem of “How can we model empathy in embodied conversational agents?", this research has attempted to begin the process of modeling and implementing levels of empathic behavior based on the theoretical foundations of empathy from various fields including philosophy, psychology, neuroscience and ethology. This thesis consisted of a series of publications that answer four research questions that were laid out in the introduction (Chapter 1). Three of the questions were focused on constructing theoretical foundations on how to model empathy in embodied conversational agents:

- **RQ1** What are the theoretical requirements/components for an empathic agent?
- **RQ2** How can a computational empathy model be described to be efficiently used by a virtual agent?
- **RQ3** How can we evaluate an empathic virtual agent?

These questions were addressed in Chapters 2, 3 and 4, respectively. Chapter 2 laid out the review of the perspectives and models on empathy from various fields with to gather requirements for empathic behavior. This chapter is intended to answer **RQ1**. Following from this background, Chapter 3 showed how these requirements could be translated into components of an empathy model to be implemented in interactive agents, inspired by the influential work by Preston and de Waal (Preston & de Waal, 2002; de Waal, 2007). The components included emotional communication, emotion regulation and cognitive mechanisms that formed a hierarchical organization to achieve levels of empathic behavior. This chapter was addressing **RQ2** by providing an example model that can be integrated into a virtual agent. As the novel field of computational empathy lacks standard measures of evaluation, Chapter 4 focused on what the evaluation methods for empathic interactive agents could look like and how it can be addressed **RQ3**. Drawing conclusions from evaluation metrics in psychology and HCI research, the chapter provided recommendations and a road map for evaluation methods in computational empathy.

After establishing theoretical foundations, Chapters 5, 6, 7 and 8 were aimed at answering the last research question:

- **RQ4** How can an empathy model be implemented in an embodied conversational agent framework?

Chapter 5 provided a detailed framework and implementation for an empathic embodied conversational agent that is capable of real-time multi-modal interaction. The chapter demonstrated in detail how the modules of the framework could process the visual and audio input from the user to generate levels of empathic behavior that were presented in Chapter 3. The implementation relies on various machine learning techniques that are used frequently in affective computing and artificial intelligence research, mainly on the perceptual module that processes the inputs from the user. Chapter 6 focuses on the automatic gesture generation system based on the emotion regulation parameters separately. This gesture generation module ensures the quality of service in the multi-channel empathic output of the agent, depending on the emotions, mood and expressivity of the agent. Both of these chapters focus on providing an automated, fluent and natural interaction that the empathic behaviors can be founded upon.

Using the framework described in Chapters 5 and 6, and the empathy model defined in Chapter 3; Chapters 7 and 8 focused on the implementation of specific empathic behaviors and their evaluation. In the scope of this thesis, we implemented the lowest level in our empathy model (emotional communication competence), some of the mid-level emotion regulation capabilities that can be associated with mood and personality and only a few high-level functionality such as a simple user model and context-based dialogue reasoning system to achieve different levels of empathic behavior. The higher levels of cognitive mechanisms such as appraisal behavior and perspective taking were not included in the implementation as these are still unsolved problems in artificial intelligence research. The main contribution of the thesis is providing a structured framework to be able to include these higher level behaviors easily for future development.

The evaluation of the empathic behaviors of the embodied conversational agent was completed following the proposed methods in Chapter 4. Chapter 7 presented the implementation of low-level empathic behavior such as mimicry and affect-matching in addition to the baseline listening gestures as well as verbal response. Results showed that, indeed, the implemented behaviors increased the perception of empathy in mimicry and affect-matching behavior as opposed to the baseline behavior. Moreover, affect-matching behaviors showed a significant increase in the perception of empathy only in the verbal feedback condition. Chapter 8 focused in detail on the empathic concern and coping behaviors that are expressed via dialogue. An example scenario of the Psychological Counseling Service intake process was used where the overarching goals of the agent were to conduct surveys and gather information about the user. Results showed that the implementation of proposed behaviors leads to an increase in the perception of empathy, as well as perceived usefulness, human-likeness and believability of the agent during the intake scenario.

Overall, the findings of this research suggested that equipping interactive agents with the hierarchical components of computational empathy that we theoretically constructed and implemented, did not only point to a significant increase in the perception of empathy but also enhanced the interaction (see Chapter 8.6 and 8.7 that shows our results and discussion). Moreover, our work showed that modeling embodied conversational agents can provide us with a testbed for examining empathic behavior and its perception, which cannot be studied efficiently in humans (see Chapter 7.5 for a discussion). Additionally, the datasets and the code generated from this dissertation are shared publicly to contribute to computational empathy research, initiate the conversations to improve the field, and enhance our understanding of empathy (see Appendix A for a link to shared code and the datasets).

Two years ago Paiva and colleagues (2017) published the first survey on the field of computational empathy, where the earliest attempts of empathic agents were seen in the early 2000s (Prendinger & Ishizuka, 2005; Brave et al., 2005). Their survey has addressed three of the six major challenges in this novel field which they described as: development of theoretically grounded models (Chapter 3), development of evaluation methods (Chapter 4) and achieving better autonomy (Chapters 5 to 8). We have also discussed the problem of context, which was also mentioned as one of the most important aspects of empathic behavior by Omdahl (Omdahl, 1995), in Chapter 8. Also, throughout this dissertation, we focused on the integration of state-of-the-art AI techniques to achieve fluency and better interaction, while aiming at theoretically-grounded and non-binary empathic behavior that we have addressed as central issues in computational empathy research (see Chapter 2).

9.2 Limitations

During both the conceptualization and implementation phases that were presented in the previous chapters, we faced some constraints that we are going to mention in this section.

9.2.1 Disagreement between Empathy Theories

There has been diverse contributions to definitions as well as models of empathy from a variety of disciplines for many decades (see Chapter 2 for an extensive overview). Even though this diversity allows for an in-depth examination of the concept, it also poses a challenge on finding a common-ground. The initial challenge we faced was to choose a foundational definition and model to base our artificial empathy framework. As a guide to our empathy model, we chose the evolutionary model from de Waal and Preston (de Waal, 2007; Preston & de Waal, 2002) that captured a variety of empathic behavior as an umbrella term (see Chapter 3). Similar to developmental theories, this model contains low-level behaviors such as mimicry and affect matching as empathic behaviors that are based on basic Perception-Action Mechanisms. Other researchers, such as Coplan (Coplan, 2011), would not agree with this broad view of empathy and argue that empathy should only contain the high-level behaviors such as self and other oriented perspective taking. According to this view, the mimicry behavior that we have focused in Chapter 7 would not be considered as empathic behavior nor the affective matching behavior without higher-level processes and a clear self-other distinction.

However, from an implementation point-of-view, we believe it is necessary to include low-level behaviors to successfully integrate empathy in artificial agents. Higher-levels of empathic behavior would require basic emotion recognition and expression mechanisms, and any kind of affect matching behavior without other-oriented evaluation to be able to reason with the input and express the emotions. Therefore, the lower-levels of capabilities are needed to be implemented for any type of higher-level behavior to function properly. Moreover, these types of responses are not only useful in the context of achieving human-like empathy, but also in terms of the naturalness and fluency of behaviors that the agent would embody. Lower-level emotional responses allow for immediate feedback to the user, thus, shortening the waiting time. An agent that would solely focus on higher-level behaviors would certainly introduce longer processing times with the current state of technology, which would impact real-time interaction.

9.2.2 Lack of Validated Evaluation Methods

Even though this novel field of computational empathy is growing rapidly, it lacks methods and standards for evaluating empathic behavior. As a part of this thesis and the research questions (RQ3), we have proposed a guideline to evaluate empathy in interactive agents (see Chapter 4). We used the main principles that were laid out in Chapter 4 for the evaluation phase of our system in Chapters 7 and 8. We used a modified version of the well-established self-report questionnaires of empathy for the primary evaluation metric on measuring perceived empathy of the agent. These self-report questionnaire were also suggested by Paiva and colleagues (Paiva et al., 2017) as a part of the EMOTE project. This allowed us to measure empathy levels quantitatively as a continuum, where most of the previous research relied on a binary categorization of empathy.

However, one major problem of these metrics is that they have not been validated in the context of human-computer interaction. One possibility to validate the adapted surveys in the context of computational empathy could be to conduct a meta-analysis over the proposed feature-based evaluations and the system-based evaluations. Another approach is to get expert opinions, as it is often done in the empathy evaluations in humans, by comparing the results of expert analysis over the agent behavior to the survey results.

Moreover, as these metrics were initially intended for general human empathy where the evaluation is based on the overall behavior, it is not clear if these metrics will be suitable for evaluations based on perceptions from a short-term interaction. During our user studies (Chapter 8) we received feedback from a few of our subjects that a number of questions in the empathy evaluation metric felt disconnected to the nature of the interaction. Subjects stated that questions such as “The agent would become irritated when someone cries" or “The agent would find it silly for people to cry out of happiness" as confusing questions for the interactions they had with the agent. Even though we could argue that these questions are directed towards evaluating the general tendency of the agent to feel empathic, there is a need for adjustment and room for improvement for these metrics to be better suited for the computational empathy research. For a more suitable evaluation of computational empathy, we think there should be a collective effort in the field to achieve this goal. An aspect of this thesis, noted in papers 3, 7 and 8, is an attempt to move the field forward in empathy and affective computing-based evaluation techniques that would help in collectively developing validated standards.

9.2.3 Limitations Due to Implementation Decisions

During the implementation phase of our model, we had to make several design choices that would force us to ignore or restrict some aspects of empathy. In the following, we will focus on those elements.

Emotion Representations

Researchers of affective computing have been using a variety of approaches to represent emotions to be used in the categorization of recognized emotions as well as expressions. In our research, we chose categorical representations of emotions as the basic representation that is shared between different modules in our implementation (see Chapter 5). During the emotion expression evaluation phase, we included the dimensional representations of P-A-D (see Chapter 6) as well as the categorical presentation, and our sentiment analysis module uses categorical as well as P-A-D representation space (see Chapter 5). However, during the integration of the modules with the rest of the framework including the empathy categories, dialogue management and behavior expression, we relied on the categorical representations of emotions. As Ekman's categories only include six basic emotions (anger, disgust, fear, happiness, sadness, surprise), it can be argued that they do not provide a rich representation format to capture the complexity of affective states of the user. See Gunes et al. (2011) for an explanation of this debate, which states that everyday interactions require more complex, non-basic and subtle affective states that may not be adequately represented by a single label. Moreover, this might restrict the array of empathic emotions that can be expressed by the agent.

However, given the novelty of this emerging field, the decision to use categorical representations were based on starting with a small set of emotions that can provide us with a starting point for comparing the basic empathic responses (see Chapter 7). Moreover, most of the state-of-the art emotion recognition methods used from affective computing research focus on a categorical representation. Nevertheless, our framework is highly modular and allows for changing the representations as well as the methods used for emotional communication capabilities. Providing our framework as open-source software, we hope that a variety of these choices can be further examined by computational empathy scholars.

Inter-personal Emotion Regulation Capabilities

As we have covered in Chapter 2, many models of empathy acknowledge the importance of intra-personal parameters that affect the empathy of a person. These parameters include factors such as similarity, liking, friendship bias and familiarity, which we included in our theoretical model in Chapter 3 as relationship related regulation factors. However, in our implementation of the regulation mechanisms that are explained in Chapter 5 and 6, we did not include these factors in our model. This is due to the specific application scenario in which the agent is desired to achieve an unbiased empathic interaction with the user that all of the users were treated equally.

Moreover, to be able to compare the within-subjects effect of the empathic behaviors quantitatively, consistent behavior was required to be presented by the agent. Although this decision might be criticized as an important distinction from the human-like empathy behavior, it can be argued that introducing a familiarity bias in a short-term human-computer interaction might not be desirable. There are examples in computational empathy research, where the agents show varying degrees of empathic behavior in agent-to-agent scenarios that can be applied within the framework we are proposing.

In Boukricha and colleague's work (2013), the virtual agents EMMA and MAX engaged in an interaction where the neutral, medium and maximum empathy levels were recognized in relation to the degree of relationship between the agents. The closer the relationship/friendship, higher degree of empathy was felt between the avatars, which seemed to be justified by the closeness of the agents. Moreover, empathy research in humans suggests that empathy between interaction partners might increase during a long-term interaction. However, as the behavior of the agent would also affect the attitude of the user, in a scenario where the agent will be acting as an assistant, a counselor, or a caregiver; it would be important to initiate and maintain the trust and positive attitude of the user with a high-level of empathic behavior. An example of this idea can be seen in the work of Brave and colleagues (2005), where the users interacted with an agent in a game environment and found the empathic agent to be more likable, trustworthy and caring to the non-empathic counterpart. Similarly, in an interaction scenario with a virtual trainer, a sustained high-level of empathic agent behavior is shown to affect the user's attitude positively towards long-term interaction (Bickmore & Picard, 2005).

Affective Speech

The main objective of this dissertation was to provide a theoretically grounded implementation of empathic behaviors in an embodied conversational agent. By doing so, our focus was not on generating new methods and algorithms for every individual mechanism, but instead creating a system that unites the existing state-of-the art methods that would be suitable for generating empathy mechanisms. Even though we have implemented applications for an emotion recognition and expression system (described in Chapter 5), an emotion regulation system (described in Chapter 6), a dialogue system (described in Chapter 8) from scratch; we have also incorporated modified versions of state-of-the-art realization methods for several individual components, which we have re-implemented for our needs. These components include the speech recognition system from Google Cloud Speech, the sentiment analysis methods used from the VADER and SOCAL sentiment analyzers, and the SmartBody character animation platform (see Chapters 5 for details). As such, the success of our system, in terms of the recognition and reasoning capabilities, depend on the accuracy of these techniques.

We incorporated these state-of-the-art and open-source tools within our framework to increase interaction success and to ensure a high-level of naturalness with our 3D character with extensive facial expressions, lip-synchronization, body gestures and gaze control in real-time. However, we still encountered some challenges due to the lack of options with better quality where the field is still evolving. One significant limitation was the Text-to-Speech (TTS) system we were using, where we were not able to control the intonation, speed and loudness of the voice of our agent. These parameters are essential in the expression of emotions, and the lack of these parameters can create a disparity between the highly expressive nature of the agent. This issue was brought to our attention during the demonstrations of our agent in multiple venues as well as our first evaluation study in Chapter 7. The inconsistency between this less expressive voice and the natural appearance of the agent with fluent facial/bodily gestures, gaze behavior and posture seem to affect the judgment of the user towards the agent. However, as training and generating more effective recognition and expression systems were beyond the scope of this work, we decided to continue using the best approach we could find with our studies.

One alternative to this could have been using voices of real humans, with the proper inflection and naturality, and using a non-dynamic system where we could only answer via a set of pre-recorded human voice utterances. However, this would not give us the desired control over generating the utterances dynamically, as the system would have to be strictly domain constrained as it would be impossible to record all possible utterances with possible emotional values in advance. Currently there are better options that uses state-of-the-art machine learning techniques to achieve a natural sounding and parameterized TTS systems. The WaveNet architecture (Van Den Oord et al., 2016; Chen et al., 2018) has recently been developed by researchers at Google which can be considered as one of the most successful alternatives to a standard TTS system. WaveNet includes SSML support for annotating pauses, change of speed, sound level, emphasis, speaking rate and pitch changes. Although these systems provide much more natural generative speech, the representation between emotions to the parameters needs to be configured separately. However, WaveNet architecture and other similar alternatives such as Amazon Polly (*Amazon Polly*, n.d.) are cloud-based systems which might introduce a considerable delay during the preparation of the sound file. Moreover, as these systems do not provide phoneme and timing information automatically, integration of these systems to a lip-synchronization tool would require implementation of a component that allows the communication between the produced audio file to the animation system.

The natural emotional intonation of a synthetic voice at real-time that can be controlled via emotional parameters would have been ideal for these studies, but its research efforts are currently underway. With the future improvements in the field, such a component can be included modularly into our system where the expression of affect by the voice that would match the quality of the rest of the generated agent behaviors.

9.3 Future Work

This thesis approaches the computational empathy research as an interdisciplinary effort that requires collaborations of multiple fields. We believe that our contributions of establishing theoretical requirements, a theoretically grounded model for computational empathy, the implementation of empathic behaviors, and providing guidelines for evaluations would lead to further enhancement of the field. By providing a modular framework and an open-sourced implementation, we aim to build further collaborations to extend our impact beyond this dissertation. Some of the envisioned improvements for our system were mentioned above in the limitations section, which can be integrated with the new advances in the field of artificial intelligence. In the following sub-sections, we will go over these and further improvements for future work.

9.3.1 Implementing Cognitive Mechanisms

As we have mentioned both in Chapter 2 while gathering the components required for empathic behaviors and in Chapter 3 during the construction of our model of empathy for computational agents, we have mentioned three hierarchical levels of components of empathy. The lowest level consisted of the emotional communication mechanisms which we have shown the implementation (Chapter 5 and 7) and the evaluation (Chapter 7). The mid-level consisted of the regulation mechanisms which we have implemented (Chapter 5, 6 and 8) and evaluated (Chapter 8). The high-level components include the cognitive mechanisms, which are the appraisal, theory-of-mind and perspective-taking abilities that include problems where most of them are not solved in the computational cognitive science community quite yet. These high-level behaviors would have required a concentrated effort that is dedicated to only one of the process where even the implementation of only the appraisal mechanisms would necessitate an equal amount of work that this dissertation has provided.

Therefore, during the conceptualization phase of the thesis goals and planned contributions, we chose to only provide the theoretical foundations for requirements (Chapters 2.4 and 3.4) and recommendations for future implementation (Chapters 3.5 and 5.5.2) of this last level of empathic behavior. We have constructed the model of the user's overall affective state and the goals of the agent combined with the context-understanding capabilities in the dialogue management system which can be used as a base-line for higher levels of cognitive mechanisms. By providing a modular and open-source implementation of the first two levels of empathic behavior, the high-level behaviors could be integrated into the system in future work with further collaborations.

9.3.2 Validated evaluation metrics

One of the main contributions of this thesis was to initiate the conversation on suitable evaluation metrics and methods of computational empathy in interactive agents. We have laid out the foundations for systematic evaluation in Chapter 4 of this thesis and used the main ideas from that work into the evaluations of our system in Chapters 7 and 8. During the evaluation phase of our implementation in Chapters 7 and 8, the proposed feature-based evaluations for the agent were not presented due to the scope of the publications. Even though the emotional repertoire of the agent were evaluated separately in Chapter 6, these evaluations as well as the emotion recognition capabilities were not tied to the overall evaluations of the agent. In Chapter 7, the accuracy of emotion recognition and expression behaviors were controlled, as well as the success of the speech feedback. The variety within the interactions during the evaluation phase of Chapter 8 was also controlled via constraining the scope of the conversation as well as the accepted responses. However, the interactions still included a variety of responses both in disclosure of information and emotional content provided by the participants. Concerning the importance of the success of these behaviors on the perception of the agent's empathic capabilities, additional examination of the accuracy of these capabilities could be added to future work.

Moreover, as explained in the limitations section of this chapter, there should be a collective effort in the field to achieve the goal of constructing evaluation metrics that can be used to assess the empathy of artificial agents. We will be presenting our work and findings in scientific venues in the upcoming months and will engage with the lead figures in empathy research to discuss possibilities. For future work, we hope to be actively involved in this effort.

9.3.3 User-related Factors

This thesis focused mainly on a theoretically-grounded implementation and evaluation of computational empathy from the user's perspective while comparing different levels of empathic behavior as it is perceived by the users. During the evaluations of the agent, we used a within-subjects design that compares the perception of different behaviors by the same subjects. Another approach that can be taken, which would also be beneficial to the empathy research, is understanding the effect of user-related factors. As the perception of empathy in artificial agents, distinct from the human behavior, can be influenced by many different factors such as age, gender, familiarity with virtual agents, as well as the empathy of the users themselves.

9.3.4 New Implementation Platforms and Scenarios

In the evaluation phase of this thesis, we chose the goal-directed scenario of psychological counseling intake process where the empathic capabilities of the agent could be observed at each step of the interaction. However, as mentioned extensively in the introduction and Chapter 2, empathic behaviors can improve the interaction in many application domains from education to the medical industry, personal companions to entertainment. During the implementation phase, we initiated collaborations with the industry to expand the application space for our empathic conversational agent. With our current industry associates Virtro (*Virtro Entertainment Inc.*, 2019) and NewPathVr (*NewPathVr*, 2019), we are in the process of expanding the use-cases for our system in language learning and medical counseling applications.

These collaborations also brought up the need of integrating our system with different environments such as VR, AR systems and mobile phones, and using a variety of appearances for the agent in terms of gender, age and other qualifications. Following this need, we have started to move towards implementing our system within the Unity3D platform and moving away from the Smartbody system. Unity3D would allow us easier integration with multiple devices for the audio-visual output of our agent such as VR/AR systems as well as personal devices such as phones or tablets. Additionally, this would expand the end-users for such a system by allowing it to be used outside of the lab setting. We have already constructed a behavior synchronization system with new characters as a part of this research effort (see Figure 9.1).

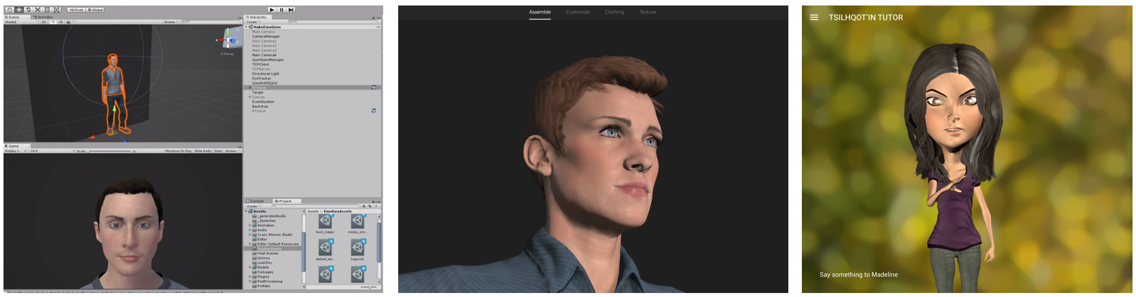

Figure 9.1: Examples of our new characters with a variety of genders and characteristics that are created in different platforms and interaction scenarios for future work.

For future work, we aim to expand these collaborations and evaluate our system in the consumer environment with a possibility of long-term interactions with the industry. The effect of realism of these characters could also be evaluated by using characters with varying levels of visual realism and behavioral realism (Bailenson et al., 2006). New techniques for photorealistic characters with better behavioral control (Feng et al., 2017), and cartoon-like characters that could embody better facial and lip-synch behavior (Y. Zhang et al., 2016; Xu, Feng, Marsella, & Shapiro, 2013) could be used in different environments to understand the effect of realism on the perception of empathic behavior. Additionally, the impact of gender and ethnicity on both feeling empathy and perceiving empathic behavior has been extensively studied in empathy research in humans. The evaluations of our system only included a Caucasian-male character in the scope of this thesis, while the effect of race and gender is controlled in a within-subjects study. For future work, we also aim to investigate the impact of different genders or genderless agents as well as variations in ethnicity.

Moreover, additional recognition functionalities can be used to extend the capabilities of the agent, both by providing additional information to the agent and changing how an interaction scenario could look like. The emotions, mood and personality of the user as well as the social context of the interaction can be recognized by using additional channels of information. The posture of the users, head gestures and hand gestures would provide valuable information about the personality of the user as well as the context and emotions of the user (DeVault et al., 2014). The personal space the user is keeping with the agent (Saberi, Bernardet, & DiPaola, 2015, 2014) as well as the gaze behavior (Nixon et al., 2018) can indicate to the social context and personality. Moreover, additional biosensing capabilities such as heart-beat, skin conductance, breathing patterns as well as EEG signals can open up new possibilities of how an interaction between an empathic agent could look like, by giving additional information on the user. These functionalities can be tied to the expression capabilities of the agent based on mood and personality that we already implemented.

9.3.5 Ethical Considerations

It is shown by many researchers that empathy, as well as rapport building behavior can lead to increased trust felt towards a system (Leite et al., 2014; Brave et al., 2005; DeVault et al., 2014; Gratch, Kang, & Wang, 2013). Although empathic behavior is often thought as related to compassionate behavior, there are no clear indication of why would empathy alone lead to moral behavior at all. In fact, concerning the increase in the trust of the agent by mere display of low-level empathic behavior such as affect matching without the actual understanding of the context could be problematic.

A possible requirement for such agents could be the disclosure of the level of expertise and capabilities that can be provided by the agent. For example, in a psychological consulting scenario, the increase in trust for the agent could result in a blind faith in the guidance that will be given by the agent (Lucas, Gratch, King, & Morency, 2014). A simple mis-recognition in the language understanding or emotion recognition capabilities of the agent could lead to an incorrect detection on the possible suicidal tendency of the user. These mistakes can lead to catastrophic consequences and the limits of the agent's behavior should always be disclosed and the severity of the cases should be checked by a human. Moreover, the amount of the information that will be kept anonymized or shared to third parties should always be disclosed.

In conclusion, equipping agents with empathic behavior is a novel research idea that should be carefully investigated in terms of the additional ethical challenges it will bring to interactive agents research.

9.4 Final Remarks

The overarching objective of this research was to provide insights on a theory-driven design approach of computational empathy. To achieve this objective, this thesis systematically addressed the problem of modeling empathy in embodied conversational agents by creating theoretical foundations, models and evaluation methods to guide the implementation. The levels of empathic behavior based on these foundations have been implemented in a state-of-the-art 3D virtual human that is capable of generating automated natural and believable gestures as well as speech coupled with the real-time multi-modal interaction abilities. These contributions provide a way to improve our interaction and the perception of empathy in users, which is shown by the evaluations. We hope our systematic approach, theoretical contributions and the open-sourced implementations with the generated data will enhance the novel field of computational empathy.

References

Adams Jr, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological science*, *14*(6), 644–647.

Amazon polly. (n.d.). `https://aws.amazon.com/polly/`. (Accessed: 2019-05-30)

Arnold, M. B. (1960). Emotion and personality.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612.

Asada, M. (2015). Towards artificial empathy. *International Journal of Social Robotics*, *7*(1), 19–33.

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, *15*(4), 359–372.

Ball, G., & Breese, J. (2000). Emotion and personality in a conversational agent. *Embodied conversational agents*, 189–219.

Barendregt, W. (2016). *Scope of the evaluations and appropriate methods.* `http://gaips.inesc-id.pt/emote/wp-content/uploads/2016/03/Month-24-D7.1.pdf`. (Accessed: 2019-05-30)

Bar-On, R., Tranel, D., Denburg, N. L., & Bechara, A. (2004). Emotional and social intelligence. *Social neuroscience: key readings*, *223*, 52.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child psychology and Psychiatry*, *38*(7), 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of developmental psychology*, *4*(2), 113–125.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with asperger syndrome or high–functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 361–374.

Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. (2013). *Understanding other minds: Perspectives from developmental social neuroscience.* Oxford University Press.

Baron-Cohen, S., & Wheelwright, S. (2003). The friendship questionnaire: An investigation of adults with asperger syndrome or high-functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *33*(5), 509–517.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *34*(2), 163–175.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251.

Baron-Cohen, S., Wheelwright, S., Jolliffe, & Therese. (1997). Is there a" language of the eyes"? evidence from normal adults, and adults with autism or asperger syndrome. *Visual cognition*, *4*(3), 311–331.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., Lawson, J., et al. (2001). Are intuitive physics and intuitive psychology independent? a test with children with asperger syndrome. *Journal of Developmental and Learning Disorders*, *5*(1), 47–78.

Barros, P. (in press). Omg empathy challenge. *IEEE international conference on automatic face and gesture recognition.*

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71–81.

Barton, K. (2018). Mumin: Multi-model inference [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=MuMIn` (R package version 1.42.1)

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bates, J., et al. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*(7), 122–125.

Batson, C. D. (2009). These things called empathy: eight related but distinct phenomena.

Batson, C. D. (2012). The empathy-altruism hypothesis: Issues and implications. *Empathy: From bench to bedside*, 41–54.

Batt-Rawden, S. A., Chisolm, M. S., Anton, B., & Flickinger, T. E. (2013). Teaching empathy to medical students: an updated, systematic review. *Academic Medicine*, *88*(8), 1171–1177.

Beale, R., & Creed, C. (2009). Affective interaction: How emotional agents affect users. *International Journal of Human - Computer Studies*, *67*(9), 755–776.

Becker, C., Prendinger, H., Ishizuka, M., & Wachsmuth, I. (2005). Evaluating affective feedback of the 3d agent max in a competitive cards game. In *International conference on affective computing and intelligent interaction* (pp. 466–473).

Bekoff, M. (2010). *The emotional lives of animals: A leading scientist explores animal joy, sorrow, and empathy—and why they matter.* New World Library.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, *3*(Feb), 1137–1155.

Bernardet, U., & DiPaola, S. (2015). A framework for exogenous and endogenous reflexive behavior in virtual characters. In *International conference on intelligent virtual agents* (pp. 132–136).

Bernardet, U., Kang, S.-h., Feng, A., DiPaola, S., & Shapiro, A. (2017). A dynamic speech breathing system for virtual characters. In *International conference on intelligent virtual agents* (pp. 43–52).

Bernardet, U., Saberi, M., & DiPaola, S. (2016). Simulink toolbox for real-time virtual character control. In *International conference on intelligent virtual agents* (pp. 456–459).

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293–327.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: analyzing text with the natural language toolkit.* O'Reilly Media, Inc.

Boukricha, H., Wachsmuth, I., Carminati, M. N., & Knoeferle, P. (2013). A computational model of empathy: Empirical evaluation. In *Affective computing and intelligent interaction (acii), 2013 humaine association conference on* (pp. 1–6).

Bower, G. H. (1981). Mood and memory. *American psychologist*, *36*(2), 129.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Bratman, M. (1987). Intention, plans, and practical reason.

Brave, S., & Nass, C. (2003). Emotion in human-computer interaction. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, 81–96.

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, *62*(2), 161–178.

Briggs, J. (2018). *Stomp messaging library for python.* Retrieved from `https://pypi.org/project/stomp.py/4.1.21/`

Broekens, J., Degroot, D., & Kosters, W. A. (2008). Formal models of appraisal: Theory, specification, and computational model. *Cognitive Systems Research*, *9*(3), 173–197.

Burleson, W., & Picard, R. W. (2004). Affective agents: Sustaining motivation to learn through failure and a state of stuck. In *Workshop on social and emotional intelligence in learning environments.*

Burleson, W., & Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. *IEEE Intelligent Systems*, *22*(4).

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, *1*(1), 18–37.

Calvo, R. A., D'Mello, S., Gratch, J., & Kappas, A. (2015). *The oxford handbook of affective computing*. Oxford Library of Psychology.

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, *31*(2), 102–107.

Campos, J. J., Walle, E. A., Dahl, A., & Main, A. (2011). Reconceptualizing emotion regulation. *Emotion Review*, *3*(1), 26–35.

Carey, K., Saltz, E., Rosenbloom, J., Micheli, M., Choi, J., & Hammer, J. (2017). Toward measuring empathy in virtual reality. In *Extended abstracts publication of the annual symposium on computer-human interaction in play* (pp. 551–559). ACM.

Cassell, J., Bickmore, T., Campbell, L., & Vilhjálmsson, H. (2000). Designing embodied conversational agents. *Embodied conversational agents*, 29–63.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000). Conversation as a System Framework: Designing Embodied Conversational Agents. In (pp. 29–63). MIT Press.

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2004). Beat: the behavior expression animation toolkit. In *Life-like characters* (pp. 163–185). Springer.

Charles, G. (1981). *Conversational organization: Interaction between speakers and hearers.* New York, Academic Press.

Chartrand, T. L., & Van Baaren, R. (2009). Human mimicry. *Advances in experimental social psychology*, *41*, 219–274.

Chen, Y., Assael, Y., Shillingford, B., Budden, D., Reed, S., Zen, H., ... others (2018). Sample efficient adaptive text-to-speech. *arXiv preprint arXiv:1809.10460*.

Chi, D., Costa, M., Zhao, L., & Badler, N. (2000). The emote model for effort and shape. In *Proceedings of the 27th annual conference on computer graphics and interactive techniques* (pp. 173–182).

Chiu, C.-C., & Marsella, S. (2011). How to train your avatar: A data driven approach to gesture generation. In *International workshop on intelligent virtual agents* (pp. 127–140).

Churchill, E. F., Cook, L., Hodgson, P., Prevost, S., & Sullivan, J. W. (2001). "may i help you?": designing embodied conversational agent allies. In *Embodied conversational agents* (pp. 64–94).

Clark, A. J. (2014). *Empathy in counseling and psychotherapy: Perspectives and practices.* Routledge.

Clavel, C., & Callejas, Z. (2016). Sentiment Analysis: From Opinion Mining to Human-Agent Interaction. *IEEE Transactions on Affective Computing*, *7*(1), 74–93. Retrieved 2018-12-05TZ, from `http://ieeexplore.ieee.org/document/7122903/` doi: 10.1109/TAFFC.2015.2444846

Clavel, C., Pelachaud, C., & Ochs, M. (2013). User's sentiment analysis in face-to-face humanagent interactions-prospects. In *Workshop on affective social signal computing, satellite of interspeech, grenoble.*

Cooper, B., Brna, P., & Martins, A. (2000). Effective affective in intelligent systems–building on evidence of empathy in teaching and learning. In *Affective interactions* (pp. 21–34). Springer.

Coplan, A. (2011). Understanding empathy: Its features and effects. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 3–18). Oxford University Press.

Coplan, A., & Goldie, P. (2011). *Empathy: Philosophical and psychological perspectives.* Oxford University Press.

Costa, P. (1994). *Revised neo personality inventory (Neo Pi-R) and neo five-factor inventory (Neo-Ffi): professional manual.* Lutz, Fl.: Psychological Assessment Resources. (OCLC: 771112626)

Costa, P. T., & MacCrae, R. R. (1992). *Revised neo personality inventory (neo pi-r) and neo five-factor inventory (neo-ffi): Professional manual.* Psychological Assessment Resources, Incorporated.

Crawford, N. C. (2014). Institutionalizing passion in world politics: Fear and empathy. *International Theory*, *6*(3), 535–557.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085.

Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain.* New York: Putnam.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

Davis, M. H. (1994). *Empathy: A social psychological approach.* Brown and Benchmark Publishers.

Davis, M. H., et al. (1980). A multidimensional approach to individual differences in empathy.

Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in cognitive sciences*, *18*(7), 337–339.

Decety, J., & Ickes, W. (2011). *The social neuroscience of empathy.* MIT Press.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., ... others (2014). Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on autonomous agents and multi-agent systems* (pp. 1061–1068).

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in cognitive sciences*, *10*(10), 435–441.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding.*

de Waal, F. B. (2007). The 'russian doll'model of empathy and imitation. *On being moved: From mirror neurons to empathy*, 35–48.

de Waal, F. B. (2010). *The age of empathy: Nature's lessons for a kinder society.* Broadway Books.

de Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498.

D'mello, S., & Graesser, A. (2012). Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 23.

D'Mello, S., Kappas, A., & Gratch, J. (2018). The affective computing approach to affect measurement. *Emotion Review*, *10*(2), 174–183.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, *23*(2), 283.

Dybkjaer, L., Bernsen, N. O., & Minker, W. (2004). Evaluation and usability of multimodal spoken language dialogue systems. *Speech Communication*, *43*(1-2), 33–54.

Eisenberg, N., & Strayer, J. (1987). *Empathy and its development.*

Ekman, P. (1972). *Emotion in the human face: guide-lines for research and an integration of findings.*

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169–200.

Ekman, P. (1993). Facial expression and emotion. *American psychologist*, *48*(4), 384.

Ekman, P. (2004). Emotional and conversational nonverbal signals. In J. M. Larrazabal & L. A. Pérez Miranda (Eds.), *Language, knowledge, and representation* (pp. 39–50). Dordrecht: Kluwer Academic Publ. (OCLC: 845698566)

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of communication*, *22*(4), 353–374.

Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (facs).*

El-Nasr, M. S., Yen, J., & Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, *3*(3), 219–257.

Eysenck, S. B., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological reports*, *43*(3_suppl), 1247–1255.

Feng, A., Rosenberg, E. S., & Shapiro, A. (2017). Just-in-time, viable, 3-d avatars from scans. *Computer Animation and Virtual Worlds*, *28*(3-4), e1769.

Feshbach, N. D., & Feshbach, S. (2009). Empathy and education. *The social neuroscience of empathy*, *85*, 98.

Froman, R. D., & Peloquin, S. M. (2001). Rethinking the use of the hogan empathy scale: A critical psychometric analysis. *The American Journal of Occupational Therapy*, *55*(5), 566–572.

Gallaher, P. E. (1992). Individual differences in nonverbal behavior: Dimensions of style. *Journal of personality and social psychology*, *63*(1), 133.

Garton, A. F., & Gringart, E. (2005). The development of a scale to measure empathy in 8-and 9-year old children. *Australian Journal of Educational & Developmental Psychology*, *5*, 17–25.

Gerdes, K. E., Segal, E. A., & Lietz, C. A. (2010). Conceptualising and measuring empathy. *British Journal of Social Work*, *40*(7), 2326–2343.

Giannakopoulos, T. (2015). pyaudioanalysis: An open-source python library for audio signal analysis. *PloS one*, *10*(12).

Gibson, J., Malandrakis, N., Romero, F., Atkins, D. C., & Narayanan, S. S. (2015). Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In *Sixteenth annual conference of the international speech communication association.*

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The cambridge mindreading (cam) face-voice battery: Testing complex emotion recognition in adults with and without asperger syndrome. *Journal of autism and developmental disorders*, *36*(2), 169–183.

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The “reading the mind in films” task: complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience,*, *1*(2), 111–123.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. (2007). The ‘reading the mind in the voice’test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of autism and developmental disorders*, *37*(6), 1096–1106.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* Oxford University Press.

Goldman, A. I. (2011). Two routes to empathy. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 31–44). Oxford University Press.

Goldman, A. I. (2012). Theory of Mind. *The Oxford Handbook of Philosophy of Cognitive Science*. Retrieved 2018-12-05TZ, from `http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195309799.001.0001/oxfordhb-9780195309799-e-17` doi: 10.1093/oxfordhb/9780195309799.013.0017

Goleman, D. (2005). *Emotional intelligence* (10th anniversary trade pbk. ed ed.). New York: Bantam Books.

Gonsior, B., Sosnowski, S., Mayer, C., Blume, J., Radig, B., Wollherr, D., & Kühnlenz, K. (2011). Improving aspects of empathy and subjective performance for hri through mirroring facial expressions. In *Ro-man, 2011 ieee* (pp. 350–356).

Google. (2014-2017). *Google cloud speech api for python.* Retrieved from `https://cloud.google.com/speech-to-text/`

Gratch, J., Kang, S.-H., & Wang, N. (2013). Using social agents to explore theories of rapport and emotional resonance. *Social Emotions in Nature and Artifact*, *181*.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5*(4), 269–306. Retrieved 2018-12-05TZ, from `http://linkinghub.elsevier.com/retrieve/pii/S1389041704000142` doi: 10.1016/j.cogsys.2004.02.002

Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent systems*, *17*(4), 54–63.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *International workshop on intelligent virtual agents* (pp. 125–138).

Gullone, E., & Taffe, J. (2012). The emotion regulation questionnaire for children and adolescents: A psychometric evaluation. *Psychological assessment*, *24*(2), 409.

Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. In *Face and gesture 2011* (pp. 827–834).

Hagendoorn, I. (2004). Some speculative hypotheses about the nature and perception of dance and choreography. *Journal of Consciousness Studies*, *11*(F0020003), 79–110.

Halpern, J. (2001). *From detached concern to empathy: humanizing medical practice.* Oxford University Press.

Hashimoto, H., & Shiomi, K. (2002). The structure of empathy in japanese adolescents: Construction and examination of an empathy scale. *Social Behavior and Personality: an international journal*, *30*(6), 593–601.

Hatfield, E., Bensman, L., Thornton, P. D., & Rapson, R. L. (2014). New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion. *Interpersona: An International Journal on Personal Relationships*, *8*(2), 159–179.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). Emotional contagion. *Current directions in psychological science*, *2*(3), 96–100.

Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2009). Measuring acceptance of an assistive social robot: a suggested toolkit. In *Ro-man 2009-the 18th ieee international symposium on robot and human interactive communication* (pp. 528–533).

Hegel, F., Spexard, T., Wrede, B., Horstmann, G., & Vogt, T. (2006, December). Playing a different imitation game: Interaction with an empathic android robot. In (pp. 56–61). IEEE.

Hess, U., & Fischer, A. (2014). Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass*, *8*(2), 45–57.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of cognitive neuroscience*, *21*(7), 1229–1243.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press. doi: 10.1017/CBO9780511805851

Hoffman, M. L. (2001). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.

Hogan, R. (1969). Development of an empathy scale. *Journal of consulting and clinical psychology*, *33*(3), 307.

Hojat, M. (2007). *Empathy in patient care: antecedents, development, measurement, and outcomes*. Springer Science & Business Media.

Hojat, M., Mangione, S., Gonnella, J. S., Nasca, T., Veloski, J. J., & Kane, G. (2001). Empathy in medical education and patient care. *Academic Medicine*, *76*(7), 669.

Hume, D. (1739). *A treatise of human nature: being an attempt to introduce the experimental method of reasoning into moral subjects*. London: printed for John Noon.

Hume, D. (1739/2009). *A treatise of human nature being an attempt to introduce the experimental method of reasoning into moral subjects*. Auckland: The Floating Press.

Iacoboni, M. (2011). Within each other: Neural mechanisms for empathy in the primate brain. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (p. 45). Oxford University Press.

Jaimes, A., & Sebe, N. (2005). Multimodal human computer interaction: A survey. In N. Sebe, M. Lew, & T. S. Huang (Eds.), *Computer vision in human-computer interaction* (pp. 1–15). Springer Berlin Heidelberg.

Jaques, P. A., & Vicari, R. M. (2007). A BDI approach to infer student's emotions in an intelligent learning environment. *Computers & Education*, *49*(2), 360–384. Retrieved 2018-12-05TZ, from `http://linkinghub.elsevier.com/retrieve/pii/S0360131505001302` doi: 10.1016/j.compedu.2005.09.002

John, O. P., & Eng, J. (2014). Three approaches to individual differences in affect regulation: Conceptualizations, measures, and findings. In *Handbook of emotion regulation* (pp. 321–345). New York, US: Guilford Press.

Jokinen, K. (2003). *Natural interaction in spoken dialogue systems*.

Jokinen, K. (2010). *Spoken dialogue systems* (Vol. 2) (No. 1). San Rafael, Calif.: Morgan & Claypool Publishers.

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International journal of human-computer studies*, *65*(8), 724–736.

Kendon, A. (1972). Some relationships between body motion and speech. *Studies in dyadic communication*, *7*(177), 90.

Kendon, A. (1994). Do gestures communicate? a review. *Research on language and social interaction*, *27*(3), 175–200.

Kendon, A. (1997). Gesture. *Annual review of anthropology*, *26*(1), 109–128.

Kipp, M., Neff, M., & Albrecht, I. (2007). An annotation scheme for conversational gestures: how to economically capture timing and form. *Language Resources and Evaluation*, *41*(3-4), 325–339.

Kipp, M., Neff, M., Kipp, K. H., & Albrecht, I. (2007). Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In *International workshop on intelligent virtual agents* (pp. 15–28).

Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, *4*(1), 15–33.

Kobsa, A., & Wahlster, W. (Eds.). (1989). *User Models in Dialog Systems.* Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved 2018-12-05TZ, from `http://link.springer.com/10.1007/978-3-642-83230-7` doi: 10.1007/978-3-642-83230-7

Kohut, H. (2011). *The search for the self: Volume 1: Selected writings of heinz kohut 1950-1978.* Karnac Books.

Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., ... Vilhjálmsson, H. (2006). Towards a common framework for multimodal generation: The behavior markup language. In *International workshop on intelligent virtual agents* (pp. 205–217).

Kopp, S., & Wachsmuth, I. (2002). Model-based animation of co-verbal gesture. In *Proceedings of computer animation 2002 (ca 2002)* (pp. 252–257).

Kroenke, K., & Spitzer, R. L. (2002). The phq-9: a new depression diagnostic and severity measure. *Psychiatric annals*, *32*(9), 509–515.

Kumano, S., Otsuka, K., Mikami, D., Matsuda, M., & Yamato, J. (2015). Analyzing interpersonal empathy via collective impressions. *IEEE Transactions on Affective Computing*, *6*(4), 324–336.

Kuppens, P., & Verduyn, P. (2015). Looking at emotion regulation through the window of emotion dynamics. *Psychological Inquiry*, *26*(1), 72–79.

Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26.

Laffey, T. J., Cox, P. A., Schmidt, J. L., Kao, S. M., & Readk, J. Y. (1988). Real-time knowledge-based systems. *AI magazine*, *9*(1), 27–27.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the empathy quotient. *Psychological medicine*, *34*(5), 911–920.

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without asperger syndrome. *Journal of autism and developmental disorders*, *34*(3), 301–310.

Lazarus, R. S. (1966). *Psychological stress and the coping process / [by] richard s. lazarus.*

Lazarus, R. S. (1991a). *Emotion and adaptation.* New York: Oxford University Press. Retrieved 2018-12-05TZ, from http://site.ebrary.com/id/10087177 (OCLC: 47007999)

Lazarus, R. S. (1991b). Progress on a cognitive-motivational-relational theory of emotion. *American psychologist*, *46*(8), 819.

Lee, J., & Marsella, S. (2006). Nonverbal behavior generator for embodied conversational agents. In *International workshop on intelligent virtual agents* (pp. 243–255).

Lefimann, N., Kopp, S., & Wachsmuth, I. (2006). Situated interaction with a virtual human perception, action, and cognition. *Situated communication*, *166*, 287.

Leiberg, S., & Anders, S. (2006). The multiple facets of empathy: a survey of theory and evidence. In S. Anders, G. Ende, M. Junghofer, J. Kissler, & D. Wildgruber (Eds.), *Understanding emotions* (Vol. 156, pp. 419 – 440). Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S0079612306560236 doi: 10.1016/S0079-6123(06)56023-6

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic Robots for Long-term Interaction: Evaluating Social Presence, Engagement and Perceived Support in Children. *International Journal of Social Robotics*, *6*(3), 329–341. Retrieved 2018-12-05TZ, from http://link.springer.com/10.1007/s12369-014-0227-1 doi: 10.1007/s12369-014-0227-1

Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., & Dolan, B. (2016). A Persona-Based Neural Conversation Model. In (pp. 994–1003). Berlin, Germany: Association for Computational Linguistics. Retrieved 2018-12-05TZ, from http://aclweb.org/anthology/P16-1094 doi: 10.18653/v1/P16-1094

Lim, A., & Okuno, H. G. (2015). A recipe for empathy. *International Journal of Social Robotics*, *7*(1), 35–49.

Lisetti, C., Amini, R., Yasavur, U., & Rishe, N. (2013). I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)*, *4*(4), 19.

Livingstone, S. R., & Russo, F. A. (2018). The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english. *PloS one*, *13*(5), e0196391.

Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, *68*(6), 386–397.

Loper, E., & Bird, S. (2002). Nltk The natural language toolkit. In *In proceedings of the acl workshop on effective tools and methodologies for teaching natural language processing and computational linguistics. philadelphia: Association for computational linguistics.*

Loyall, A. B. (1997). *Believable agents: Building interactive personalities.* (Tech. Rep.). Carnegie-Mellon Uni Pittsburg PA.

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 ieee computer society conference on computer vision and pattern recognition-workshops* (pp. 94–101).

Maatman, R., Gratch, J., & Marsella, S. (2005). Natural behavior of a listening agent. In *International workshop on intelligent virtual agents* (pp. 25–36).

Mairesse, F., & Walker, M. A. (2010). Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, *20*(3), 227–278. Retrieved 2018-12-05TZ, from `http://link.springer.com/10.1007/s11257-010-9076-2` doi: 10.1007/s11257-010-9076-2

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60).

Marsella, S., & Gratch, J. (2003). Modeling coping behavior in virtual humans: don't worry, be happy. In *Proceedings of the second international joint conference on autonomous agents and multiagent systems* (pp. 313–320).

Marsella, S., & Gratch, J. (2014, November). Computationally modeling human emotion. *Commun. ACM*, *57*(12), 56–67. Retrieved from `http://doi.acm.org.proxy.lib.sfu.ca/10.1145/2631912` doi: 10.1145/2631912

Marsella, S. C., & Gratch, J. (2009). Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, *10*(1), 70–90.

Marsella, S. C., Johnson, W. L., & LaBore, C. (2000). Interactive pedagogical drama. In *Proceedings of the fourth international conference on autonomous agents* (pp. 301–308).

Matsuyama, Y., Bhardwaj, A., Zhao, R., Romeo, O., Akoju, S., & Cassell, J. (2016). Socially-aware animated intelligent personal assistant agent. In *Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue* (pp. 224–227).

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. e. (2016). Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the 2016 chi conference extended abstracts on human factors in computing systems* (pp. 3723–3726).

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago press.

McNeill, D. (2006). Gesture: a psycholinguistic approach. *The encyclopedia of language and linguistics*, 58–66.

McNeill, D. (2008). *Gesture and thought.* University of Chicago press.

McQuiggan, S. W., & Lester, J. C. (2007). Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies*, *65*(4), 348–360.

McQuiggan, S. W., Robison, J. L., Phillips, R., & Lester, J. C. (2008). Modeling parallel and reactive empathy in virtual agents: An inductive approach. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems-volume 1* (pp. 167–174).

McTear, M. F. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)*, *34*(1), 90–169.

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

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy 1. *Journal of personality*, *40*(4), 525–543.

Mercer, S. W., Maxwell, M., Heaney, D., & Watt, G. (2004). The consultation and relational empathy (care) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family practice*, *21*(6), 699–705.

Mercer, S. W., & Reynolds, W. J. (2002). Empathy and quality of care. *Br J Gen Pract*, *52*(Suppl), S9–12.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Minsky, M. (1991). Society of mind: a response to four reviews. *Artificial Intelligence*, *48*(3), 371–396.

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines*, *19*(3), 345.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013, June). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the seventh international workshop on semantic evaluation exercises (semeval-2013).* Atlanta, Georgia, USA.

Moller, S., Engelbrecht, K.-P., Kuhnel, C., Wechsung, I., & Weiss, B. (2009). A taxonomy of quality of service and quality of experience of multimodal human-machine interaction. In *2009 international workshop on quality of multimedia experience* (pp. 7–12).

Moridis, C. N., & Economides, A. A. (2012). Affective learning: Empathetic agents with emotional facial and tone of voice expressions. *IEEE Transactions on Affective Computing*, *3*(3), 260–272.

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of substance abuse treatment*, *28*(1), 19–26.

Moyers, T. B., Rowell, L. N., Manuel, J. K., Ernst, D., & Houck, J. M. (2016). The motivational interviewing treatment integrity code (miti 4): Rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment*, *65*, 36–42.

Müller, B. C., Van Leeuwen, M. L., Van Baaren, R. B., Bekkering, H., & Dijksterhuis, A. (2013). Empathy is a beautiful thing: Empathy predicts imitation only for attractive others. *Scandinavian journal of psychology*, *54*(5), 401–406.

Neumann, D. L., Chan, R. C., Boyle, G. J., Wang, Y., & Rae Westbury, H. (2015). Measures of empathy: Self-report, behavioral, and neuroscientific approaches. In *Measures of personality and social psychological constructs* (pp. 257–289).

Newpathvr. (2019). `https://www.newpathvr.com/`. (Accessed: 2019-05-30)

Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C. (2009). Greta: an interactive expressive eca system. In *Proceedings of the 8th international conference on autonomous agents and multiagent systems-volume 2* (pp. 1399–1400).

Nixon, M., DiPaola, S., & Bernardet, U. (2018). An eye gaze model for controlling the display of social status in believable virtual humans. In *2018 ieee conference on computational intelligence and games (cig)* (pp. 1–8).

Nowak, M. (2011). The complicated history of einfühlung. *Argument Biannual Philosophical Journal*, *1*(2), 301–326.

Ochs, M., Sadek, D., & Pelachaud, C. (2012). A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems*, *24*(3), 410–440.

Ojha, S., & Williams, M.-A. (2017). Emotional appraisal: A computational perspective. In *Fifth annual conference on advances in cognitive systems.*

Omdahl, B. L. (1995). Cognitive appraisal, emotion, and empathy.

Omdahl, B. L. (2014). *Cognitive appraisal, emotion, and empathy.* Psychology Press.

Omid Alemi and Ozge Nilay Yalcin. (n.d.). *Pymo.* Retrieved from `https://github.com/onyalcin/PyMO`

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions.* Cambridge university press.

Osgood, C. E., May, W. H., Miron, M. S., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning* (Vol. 1). University of Illinois Press.

Oviatt, S., & Cohen, P. (2000). Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, *43*(3), 45–53.

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(3), 11.

Paiva, A., Leite, I., & Ribeiro, T. (2014). Emotion modeling for social robots. *The Oxford handbook of affective computing*, 296–308.

Pantic, M., Sebe, N., Cohn, J. F., & Huang, T. (2005). Affective multimodal human-computer interaction. In *Proceedings of the 13th annual acm international conference on multimedia* (pp. 669–676).

Partala, T., & Surakka, V. (2004). The effects of affective interventions in human–computer interaction. *Interacting with computers*, *16*(2), 295–309.

Pecune, F. (2013). Toward a Computational Model of Social Relations for Artificial Companions. In (pp. 677–682). Geneva, Switzerland: IEEE. Retrieved 2018-12-05TZ, from `http://ieeexplore.ieee.org/document/6681509/` doi: 10.1109/ACII.2013.118

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pelachaud, C. (2009, Dec). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3539–3548. doi: 10.1098/rstb.2009.0186

Pelachaud, C. (2015). Greta: an interactive expressive embodied conversational agent. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems* (pp. 5–5).

Pentlan, A. (2005). Socially aware, computation and communication. *Computer*, *38*(3), 33–40. Retrieved 2018-12-05TZ, from `http://ieeexplore.ieee.org/document/1413116/` doi: 10.1109/MC.2005.104

Picard, R. W. (1997). *Affective computing* (First paperback edition ed.). Cambridge, Massachusetts London: The MIT Press. (OCLC: 247967780)

Picard, R. W. (2014). The promise of affective computing. *The Oxford handbook of affective computing*, *11*.

Picard, R. W., et al. (1995). Affective computing.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02). Retrieved 2018-12-05TZ, from `http://www.journals.cambridge.org/abstract_S0140525X04000056` doi: 10.1017/S0140525X04000056

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., & De Carolis, B. (2005). Greta. a believable embodied conversational agent. In *Multimodal intelligent information presentation* (pp. 3–25). Springer.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125.

Poria, S., Cambria, E., Howard, N., Huang, G.-B., & Hussain, A. (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*, *174*, 50–59.

Preece, D. A., Becerra, R., Robinson, K., Dandy, J., & Allan, A. (2018). Measuring emotion regulation ability across negative and positive emotions: The perth emotion regulation competency inventory. *Personality and Individual Differences*, *135*, 229–241.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(4), 515–526.

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users'affective states. *Applied Artificial Intelligence*, *19*(3-4), 267–285.

Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International journal of human-computer studies*, *62*(2), 231–245.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, *25*(1), 1–20.

Prinz, J. (2011). Is empathy necessary for morality. *Empathy: Philosophical and psychological perspectives*, *1*, 211–229.

Prochazkova, E., & Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience & Biobehavioral Reviews*, *80*, 99 - 114. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0149763416306704` doi: https://doi.org/10.1016/j.neubiorev.2017.05.013

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research*, *19*(5), e151.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a bdi-architecture. *KR*, *91*, 473–484.

Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2018). I know the feeling: Learning to converse with empathy. *arXiv preprint arXiv:1811.00207*.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press.

Reid, C., Davis, H., Horlin, C., Anderson, M., Baughman, N., & Campbell, C. (2013). The kids' empathic development scale (keds): A multi-dimensional measure of empathy in primary school-aged children. *British Journal of Developmental Psychology*, *31*(2), 231–256.

Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The qcae: A questionnaire of cognitive and affective empathy. *Journal of personality assessment*, *93*(1), 84–95.

Reynolds, W. J. (2017). *The measurement and development of empathy in nursing.* Routledge.

Riek, L. D., Paul, P. C., & Robinson, P. (2010). When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, *3*(1-2), 99–108.

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., & Robinson, P. (2009). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th acm/ieee international conference on human robot interaction* (pp. 245–246).

Rizzolatti, G., & Fabbri-Destro, M. (2010). Mirror neurons: from discovery to autism. *Experimental brain research*, *200*(3-4), 223–237.

Rodrigues, S. H., Mascarenhas, S., Dias, J., & Paiva, A. (2015). A process model of empathy for virtual agents. *Interacting with Computers*, *27*(4), 371–391.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, *10*(3), 241–278.

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory. *Appraisal processes in emotion: Theory, methods, research*, 3–19.

Rosenthal-Von Der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., ... Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fmri. *Computers in Human Behavior*, *33*, 201–212.

Roy-Charland, A., Perron, M., Beaudry, O., & Eady, K. (2014). Confusion of fear and surprise: A test of the perceptual-attentional limitation hypothesis with eye movement monitoring. *Cognition and Emotion*, *28*(7), 1214–1222.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual review of psychology*, *54*(1), 329–349.

Ruttkay, Z., Dormann, C., & Noot, H. (2006). Evaluating ecas-what, how and why? In *Dagstuhl seminar proceedings.*

Saberi, M., Bernardet, U., & DiPaola, S. (2014). An architecture for personality-based, nonverbal behavior in affective virtual humanoid character. *Procedia Computer Science*, *41*, 204–211.

Saberi, M., Bernardet, U., & DiPaola, S. (2015). Model of personality-based, nonverbal behavior in affective virtual humanoid character. In *Proceedings of the 2015 acm on international conference on multimodal interaction* (pp. 371–372).

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*(3), 185–211.

Scherer, K. R. (1982). Emotion as a process: Function, origin and regulation. *Social sciences information*, *21*(4–5), 555–570.

Scherer, K. R. (2001). *Appraisal considered as a process of multilevel sequential checking* (K. R. Scherer, A. Schorr, & T. Johnstone, Eds.). Oxford University Press.

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social science information*, *44*(4), 695–729.

Scherer, K. R. (2007). Componential emotion theory can inform models of emotional competence. In *The science of emotional intelligence : knowns and unknowns / edited by gerald matthews, moshe zeidner, and richard d. roberts.* (pp. 101–126). Oxford University Press.

Scherer, K. R. (2010a). Emotion and emotional competence: conceptual and theoretical issues for modelling agents. *Blueprint for affective computing*, 3–20.

Scherer, K. R. (2010b). The component process model: Architecture for a comprehensive computational model of emergent emotion. *Blueprint for affective computing: A sourcebook*, 47–70.

Scherer, K. R., Bänziger, T., & Roesch, E. (2010). *A blueprint for affective computing: A sourcebook and manual.* Oxford University Press.

Schroder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., … others (2012). Building autonomous sensitive artificial listeners. *IEEE Transactions on Affective Computing*, *3*(2), 165–183.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Serón, F., Baldassarri, S., & Cerezo, E. (2006). Maxineppt: using 3d virtual characters for natural interaction. In *Proc. 2nd international workshop on ubiquitous computing & ambient intelligence* (pp. 241–250).

Shapiro, A. (2011). Building a character animation system. In *International conference on motion in games* (pp. 98–109).

Shioiri, T., Someya, T., Helmeste, D., & Tang, S. W. (1999). Misinterpretation of facial expression: A cross-cultural study. *Psychiatry and Clinical Neurosciences*, *53*(1), 45–50.

Singer, T., & Lamm, C. (2009, March). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *11561*(1), 81–96.

Skowron, M. (2010). Affect listeners: Acquisition of affective states by means of conversational systems. In *Development of multimodal interfaces: Active listening and synchrony* (pp. 169–181). Springer.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Lee, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PloS one*, *8*(1), e52766.

Slote, M. (2007). *The ethics of care and empathy.* Routledge.

Smith, A. (1761). *The theory of moral sentiments / by adam smith.* (2nd ed. ed.).

Smith, A. (1959). *The theory of moral sentiments.* London: Printed for A. Millar and A. Kincaid and J. Bell.

Smith, C., Crook, N., Dobnik, S., Charlton, D., Boye, J., Pulman, S., … others (2011). Interaction strategies for an affective conversational agent. *Presence: Teleoperators and Virtual Environments*, *20*(5), 395–411.

Smith, C. A., & Kirby, L. D. (2000). Consequences require antecedents: Toward a process model of emotion elicitation.

Smith, M. (2011). Empathy, expansionism, and the extended mind. *Empathy: Philosophical and psychological perspectives*, 99–117.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.-F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3–14.

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of personality assessment*, *91*(1), 62–71.

Steinmetz, R. (1996). Human perception of jitter and media synchronization. *IEEE Journal on Selected Areas in Communications*, *14*(1), 61–72.

Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science*, *19*(10), 984.

Stueber, K. R. (2006). *Rediscovering empathy : agency, folk psychology, and the human sciences.*

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267–307.

Tan, Z. X., Goel, A., Nguyen, T.-S., & Ong, D. C. (2018, December). A multimodal lstm for predicting listener empathic responses over time. *arXiv e-prints*, arXiv:1812.04891.

Tavassoli, T., Hoekstra, R. A., & Baron-Cohen, S. (2014). The sensory perception quotient (spq): development and validation of a new sensory questionnaire for adults with and without autism. *Molecular autism*, *5*(1), 29.

Thiebaux, M., Marsella, S., Marshall, A. N., & Kallmann, M. (2008). Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems-volume 1* (pp. 151–158).

Thórisson, K. R. (2002). Natural turn-taking needs no manual: Computational theory and model, from perception to action. In *Multimodality in language and speech systems* (pp. 173–207). Springer.

Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought-processes.* Macmillan.

Traum, D. (2017). Computational approaches to dialogue. *The Routledge Handbook of Language and Dialogue*, 143.

Traum, D., DeVault, D., Lee, J., Wang, Z., & Marsella, S. (2012). Incremental dialogue understanding and feedback for multiparty, multimodal conversation. In *International conference on intelligent virtual agents* (pp. 275–288).

Traum, D. R. (1993). Mental state in the trains-92 dialogue manager. In *Working notes aaai spring symposium on reasoning about mental states: Formal theories and applications* (pp. 143–149).

Turk, M., & Robertson, G. (2000). Perceptual user interfaces. *Communications of the ACM*, *43*(3), 32–32.

Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *SSW*, *125*.

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, *5*(3), 273–291.

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schroeder, M. (2012). Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transactions on Affective Computing*, *3*(1), 69–87. Retrieved 2018-12-05TZ, from `http://ieeexplore.ieee.org/document/5989788/` doi: 10.1109/T-AFFC.2011.27

Virtro entertainment inc. (2019). `https://www.virtro.ca/`. (Accessed: 2019-05-30)

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209 - 232. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0167639313001295` doi: https://doi.org/10.1016/j.specom.2013.09.008

Wallbott, H. G. (1998). Bodily expression of emotion. *European journal of social psychology*, *28*(6), 879–896.

Wang, F.-Y., Carley, K. M., Zeng, D., & Mao, W. (2007). Social Computing: From Social Informatics to Social Intelligence. *IEEE Intelligent Systems*, *22*(2), 79–83. Retrieved 2018-12-05TZ, from `http://ieeexplore.ieee.org/document/4136863/` doi: 10.1109/MIS.2007.41

Westbury, H. R., & Neumann, D. L. (2008). Empathy-related responses to moving film stimuli depicting human and non-human animal targets in negative circumstances. *Biological psychology*, *78*(1), 66–74.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from `http://ggplot2.org`

Wilks, Y., Catizone, R., Worgan, S., & Turunen, M. (2011). Some background on dialogue management and conversational speech for dialogue systems. *Computer Speech & Language*, *25*(2), 128–139. Retrieved 2018-12-05TZ, from `http://linkinghub.elsevier.com/retrieve/pii/S0885230810000197` doi: 10.1016/j.csl.2010.03.001

Wisp, L. (1987). History of the concept of empathy. *Empathy and its development*, 17–37.

Wood, K. (2016). Kinesthetic empathy: Conditions for viewing. Oxford University Press.

Wu, J., Ghosh, S., Chollet, M., Ly, S., Mozgai, S., & Scherer, S. (2018). Nadia: Neural network driven virtual human conversation agents. In *Proceedings of the 18th international conference on intelligent virtual agents* (pp. 173–178).

Xiao, B., Bone, D., Segbroeck, M. V., Imel, Z. E., Atkins, D. C., Georgiou, P. G., & Narayanan, S. S. (2014). Modeling therapist empathy through prosody in drug addiction counseling. In *Fifteenth annual conference of the international speech communication association.*

Xu, Y., Feng, A. W., Marsella, S., & Shapiro, A. (2013). A practical and configurable lip sync method for games. In *Proceedings of motion on games* (pp. 131–140).

Yalçin, Ö. N. (2018). Modeling empathy in embodied conversational agents. In *Proceedings of the 2018 on international conference on multimodal interaction* (pp. 546–550).

Yalçın, Ö. N. (2019). Evaluating empathy in artificial agents. In *2019 international conference on affective computing and intelligent interaction (acii).*

Yalçın, Ö. N. (in press). Empathy framework for embodied conversational agents. *Cognitive Systems Research Journal.*

Yalçın, Ö. N., & DiPaola, S. (2018). A computational model of empathy for interactive agents. *Biologically Inspired Cognitive Architectures*, *26*, 20–25.

Yalçın, Ö. N., & DiPaola, S. (2019). Evaluating levels of emotional contagion with an embodied conversational agent. In *Proceedings of the 41st annual conference of the cognitive science society.*

Yalçın, Ö. N., & DiPaola, S. (2020). M-path: A conversational system for the empathic virtual agent. In A. V. Samsonovich (Ed.), *Biologically inspired cognitive architectures 2019* (pp. 597–607). Cham: Springer International Publishing.

Yngve, V. H. (1970). On getting a word in edgewise. In *Chicago linguistics society, 6th meeting, 1970* (pp. 567–578).

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, *31*(1), 39–58.

Zhang, A. (2017). *Python speechrecognition library.* Retrieved from `https://github.com/Uberi/speech_recognition#readme.`

Zhang, Y., Dong, W., Ma, C., Mei, X., Li, K., Huang, F., ... Deussen, O. (2016). Data-driven synthesis of cartoon faces using different styles. *IEEE Transactions on image processing*, *26*(1), 464–478.

Zhou, L., Gao, J., Li, D., & Shum, H. (2018). The design and implementation of xiaoice, an empathetic social chatbot. *CoRR*, *abs/1812.08989*. Retrieved from `http://arxiv.org/abs/1812.08989`

Appendix A

Code

The code generated for this research is shared in a public dataset with the following link: github.com/onyalcin/M-PATH

Appendix B

Datasets

The datasets used and generated in the scope of this research is shared at:
https://doi.org/10.48410/zr57-m463

Appendix C

Evaluation Results

The details of the evaluation results for the studies conducted in Chapter 7 and 8 will be shown in this section. Details about the datasets and evaluation methods are shared at: https://doi.org/10.48410/zr57-m463.

C.1 Evaluation Results for Chapter 7

C.1.1 Study 1 Results

In order to evaluate the perception of empathic behaviors we compared the listening behavior of the agent in backchannel, mimicry and affective matching conditions. For our study, we used within subjects design where three conditions of agent behavior are shown to the same subject for the evaluation. The conditions are baseline backchanneling behavior, mimicry with backchanneling and affective matching with backchanneling during only the listening act. We used three emotional stories told by the same person, which displays three different emotions as the main theme: joy, sadness and anger. Each video starts with a neutral remark, that is followed by the emotional story. The experiment counterbalanced on the order of the type of interaction (backchannel, mimicry, affect matching), and the order of type of emotional story (angry, sad, happy). 36 (6x6) different conditions.

In the evaluation of the first study, Mauchly's Test of Sphericity indicated that sphericity had not been violated, $X2(2) = 1.748$, $p = .417$. A one-way repeated measures ANOVA was conducted to compare the effect of (IV) level of emotional contagion behavior on (DV) the perception of empathy in backchanneling, mimicry, and affective matching conditions. The results showed that perceived empathy is significantly effected by the type of listening feedback $F(2, 70)= 16.721$ $p < .0001$, 95%CI (see Figure 7.3). Pairwise comparisons showed backchannel feedback only (M=-5.47, SD=12.45) is perceived to have significantly lower empathy than both mimicry ($p < .001$) and affective matching ($p < .0001$). However, listening behavior with mimicry (M=5.16, SD=10.64) and affective matching (M=8.22, SD=13.72) did not have any significant difference ($p = .18$).

Source		Sum of Squares	df	Mean Square	F	p	Partial η^2
empathy_level	Sphericity Assumed	3720.722	2	1860.361	16.721	.000	.323
	Greenhouse-Geisser	3720.722	1.809	2057.265	16.721	000	.323
	Huynh-Feldt	3720.722	1.901	1956.876	16.721	000	.323
	Lower-bound	3720.722	1.000	3720.722	16.721	000	.323
Error (empathy_level)	Sphericity Assumed	7787.944	70	111.256			
	Greenhouse-Geisser	7787.944	63.300	123.032			
	Huynh-Feldt	7787.944	66.548	117.028			
	Lower-bound	7787.944	35.000	222.513			

Figure C.2: Within subjects ANOVA results for Study 1

			95% Confidence Interval	
empathy_level	Mean	Std. Error	Lower Bound	Upper Bound
Backchannel	-5.472	2.077	-9.688	-1.257
Mimicry	5.167	1.774	1.565	8.768
Affect Matching	8.222	2.287	3.579	12.865

Figure C.1: Descriptive statistics for the conditions

(I) empathy_level	(J) empathy _level	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference Upper Bound
1 backchannel	2	-10.639*	2.316	.000	-16.463	-4.815
	3	-13.694*	2.860	.000	-20.887	-6.502
2 mimicry	1	10.639*	2.316	.000	4.815	16.463
	3	-3.056	2.236	.186	-8.677	2.566
3 affect matching	1	13.694*	2.860	.000	6.502	20.887
	2	3.056	2.236	.186	-2.566	8.677

Figure C.3: Pairwise comparisons for the effects

C.1.2 Study 3 Results

The third experiment focused on more complex emotional stories, where the human actor will talk about two scenarios mentioning a dog and a plant. In the dog scenario, the actor will go through excitement, disgust, worry and happiness emotions while mentioning a story about their new pet dog. In the plant scenario, the actor will go through neutral, surprise, worry and happiness emotions while mentioning a story about their friend's plant. The listening behavior of the agent will be matching the emotions both in mimicry and affective matching conditions. Similar to the second study, mimicry condition will result in a generic verbal response from the agent while affective matching condition will give an emotionally charged feedback due to emotional representation. The third experiment is also counterbalanced on the order of the type of interaction (mimicry, affect matching), and the order of the type of emotional story (dog, plant). 4 (2x2) different conditions presented to the subjects. Both experiments followed the same procedure as the first study.

One-way repeated measures ANOVA was conducted to compare the effect of (IV) level of emotional contagion behavior on (DV) the perception of empathy in mimicry, and affective matching conditions during the interaction with complex emotional behavior. The results showed that perceived empathy is significantly different between mimicry (M=0.75, SD=10.45) and affect matching (M=7.21, SD=9.98) conditions $F(1, 23) = 7.731$, $p = .011$ (see Figure 7.4).

factor1	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
mimicry	.750	2.133	-3.663	5.163
Affect matching	7.208	2.039	2.991	11.425

Figure C.4: Descriptive statistics for Study 3 parameters

Source		Sum of Squares	df	Mean Square	F	Sig.	Partial η^2	Observed Power[a]
Empathy level	Sphericity Assumed	500.521	1	500.521	7.731	.011	.252	.759
	Greenhouse-Geisser	500.521	1.000	500.521	7.731	.011	.252	.759
	Huynh-Feldt	500.521	1.000	500.521	7.731	.011	.252	.759
	Lower-bound	500.521	1.000	500.521	7.731	.011	.252	.759
Error (empathy level)	Sphericity Assumed	1488.979	23	64.738				
	Greenhouse-Geisser	1488.979	23.000	64.738				
	Huynh-Feldt	1488.979	23.000	64.738				
	Lower-bound	1488.979	23.000	64.738				

Figure C.5: Within subjects ANOVA results for Study 3

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
mimicry	Aff. match	-6.458*	2.323	.011	-11.263	-1.653
Affect match	mimicry	6.458*	2.323	.011	1.653	11.263

Figure C.6: Pairwise comparisons for the effects

C.2 Evaluation Results for Chapter 8

We evaluated the empathic screening agent to its non-empathic counterpart in a text-based interaction environment. The main difference between these two agents is their responsiveness to the emotional utterances of the user. Our hypothesis is the empathic version of the conversational agent would be perceived as more empathic, which would in turn have a positive effect on the attitude towards the interaction. A total of 16 users (10 Female, 6 Male) completed the study that were between the ages 20 and 39 (M=26.65, SD=7.74). Because we were focusing on the screening process for the student consultation service at Simon Fraser University (SFU), we chose undergraduate and graduate students in SFU. Participation in this study was voluntary and was based on open invitations to a large group of students at SFU via online communication.

From 16 users, only one encountered an unsuccessful interaction for both of the agents, where the goal of conducting the survey was not reached. None of the user responses were excluded from the final analysis of the results. All analysis and plotting are done using linear mixed models on R (R Core Team, 2018) with lme4 (D. Bates et al., 2015) package.

We performed a linear mixed effects analysis of the relationship between the perception of empathy and system type (empathic vs. non-empathic). As fixed effects, we entered the subjects into the model. Results show that perceived empathy is significantly higher in the empathic agent, relative to the non-empathic agent condition ($p = .02$).

We also examined the attitude towards the interaction. Results showed that the system type condition (empathic vs. non-empathic) significantly effects the perceived usefulness of the agent ($p = .05$). The empathic agent is found more human-like ($p < .01$) and preferred more to a human agent ($p < .01$), than the non-empathic agent. The preference of the agent over the pen-and-paper based screening process was not significantly different ($p = .2$), but high in both cases. Moreover, the results showed the system type does not have an effect on trust towards the system ($p = .41$). Figures below show the tables for the results with effect sizes.

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η^2	Observed Power[a]
Empathy level	162.0	1	162.000	6.429	.0228	.300	.713
Error (empathy level)	378.0	15	25.200				

Figure C.7: Effect of Regular and Empathic agent conditions on perceived empathy

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η^2	Observed Power[a]
Empathy level	2.0	1	2.000	4.286	.050	.222	.541
Error (empathy level)	7.0	15	0.467				

Figure C.8: Effect of Regular and Empathic agent conditions on Usefulness

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η²	Observed Power[a]
Empathy level	4.50	1	4.500	10.385	.006	.409	.893
Error (empathy level)	6.5	15	0.433				

Figure C.9: Effect of Regular and Empathic agent conditions on Human-likeness

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η²	Observed Power[a]
Empathy level	2.0	1	2.000	5	.041	.250	.605
Error (empathy level)	6.0	15	0.400				

Figure C.10: Effect of Regular and Empathic agent conditions on Believability

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η²	Observed Power[a]
Empathy level	1.125	1	1.125	9	.009	.375	.847
Error (empathy level)	1.875	15	0.125				

Figure C.11: Effect of Regular and Empathic agent conditions on Preference over human

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η²	Observed Power[a]
Empathy level	0.5	1	0.500	1.364	.261	.083	.214
Error (empathy level)	5.5	15	0.367				

Figure C.12: Effect of Regular and Empathic agent conditions on Preference over paper based survey

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial η²	Observed Power[a]
Empathy level	2.0	1	2.000	0.714	.411	.045	.135
Error (empathy level)	42.0	15	2.800				

Figure C.13: Effect of Regular and Empathic agent conditions on Trust