UNIVERSITÄT
DES
SAARLANDES

Saarland University

Faculty for Human and Business Sciences

Department of Psychology

Affective Reactions towards Socially Interactive Agents and their Computational Modeling

Dissertation
zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Fakultät für Empirische Humanwissenschaften und Wirtschaftswissenschaft
der Universität des Saarlandes

von
Tanja Schneeberger

Saarbrücken,
2023

Dekan:	Prof. Dr. Peter Loos
Berichterstattende:	Prof. Dr. Cornelius König Prof. Dr. Birgit Lugrin
Tag des Kolloquiums:	06.07.2023

Zusammenfassung

In den letzten 30 Jahren haben Forschende menschliche Reaktionen auf Maschinen untersucht und dabei das "Computer sind soziale Akteure"-Paradigma genutzt, in dem Reaktionen auf Computer mit denen auf Menschen verglichen werden. In den letzten 30 Jahren hat sich ebenfalls die Technologie weiterentwickelt, was zu einer enormen Veränderung der Computerschnittstellen und der Entwicklung von sozial interaktiven Agenten geführt hat. Dies wirft Fragen zu menschlichen Reaktionen auf sozial interaktive Agenten auf. Um diese Fragen zu beantworten, ist Wissen aus mehreren Disziplinen erforderlich, weshalb diese interdisziplinäre Dissertation innerhalb der Psychologie und Informatik angesiedelt ist. Sie zielt darauf ab, affektive Reaktionen auf sozial interaktive Agenten zu untersuchen und zu erforschen, wie diese computational modelliert werden können. Nach einer allgemeinen Einführung in das Thema gibt diese Arbeit daher, erstens, einen Überblick über das Agentensystem, das in der Arbeit verwendet wird. Zweitens wird eine Studie vorgestellt, in der eine menschliche und eine virtuelle Jobinterviewerin miteinander verglichen werden, wobei sich zeigt, dass beide Interviewerinnen bei den Versuchsteilnehmenden Schamgefühle in gleichem Maße auslösen. Drittens wird eine Studie berichtet, in der Gehorsam gegenüber sozial interaktiven Agenten untersucht wird. Die Ergebnisse deuten darauf hin, dass Versuchsteilnehmende sowohl menschlichen als auch virtuellen Anleiterinnen ähnlich gehorchen. Darüber hinaus werden durch beide Instruktorinnen gleiche Maße von Stress und Scham hervorgerufen. Viertens wird ein Biofeedback-Stressmanagementtraining mit einer sozial interaktiven Agentin vorgestellt. Die Studie zeigt, dass die virtuelle Trainerin Techniken zur Bewältigung von emotional herausfordernden sozialen Situationen vermitteln kann. Fünftens wird MARSSI, ein computergestütztes Modell des Nutzeraffekts, vorgestellt. Die Evaluation des Modells zeigt, dass es möglich ist, Sequenzen von sozialen Signalen mit affektiven Reaktionen unter Berücksichtigung von Emotionsregulationsprozessen in Beziehung zu setzen. Als letztes wird die DEEP-Methode als Ausgangspunkt für eine tiefer gehende computergestützte Modellierung von internen Emotionen vorgestellt. Die Methode kombiniert soziale Signale, verbalisierte Introspektion, Kontextinformationen und theoriegeleitetes Wissen. Eine beispielhafte Anwendung auf die Emotion Scham und ein schematisches dynamisches Bayes'sches Netz zu deren Modellierung werden dargestellt. Insgesamt liefert diese Arbeit Hinweise darauf, dass menschliche Reaktionen auf sozial interaktive Agenten den Reaktionen auf Menschen sehr ähnlich sind und dass es möglich ist diese menschlichen Reaktion computational zu modellieren.

Abstract

Over the past 30 years, researchers have studied human reactions towards machines applying the Computers Are Social Actors paradigm, which contrasts reactions towards computers with reactions towards humans. The last 30 years have also seen improvements in technology that have led to tremendous changes in computer interfaces and the development of Socially Interactive Agents. This raises the question of how humans react to Socially Interactive Agents. To answer these questions, knowledge from several disciplines is required, which is why this interdisciplinary dissertation is positioned within psychology and computer science. It aims to investigate affective reactions to Socially Interactive Agents and how these can be modeled computationally. Therefore, after a general introduction and background, this thesis first provides an overview of the Socially Interactive Agent system used in this work. Second, it presents a study comparing a human and a virtual job interviewer, which shows that both interviewers induce shame in participants to the same extent. Thirdly, it reports on a study investigating obedience towards Socially Interactive Agents. The results indicate that participants obey human and virtual instructors in similar ways. Furthermore, both types of instructors evoke feelings of stress and shame to the same extent. Fourth, a stress management training using biofeedback with a Socially Interactive Agent is presented. The study shows that a virtual trainer can teach coping techniques for emotionally challenging social situations. Fifth, it introduces MARSSI, a computational model of user affect. The evaluation of the model shows that it is possible to relate sequences of social signals to affective reactions, taking into account emotion regulation processes. Finally, the Deep method is proposed as a starting point for deeper computational modeling of internal emotions. The method combines social signals, verbalized introspection information, context information, and theory-driven knowledge. An exemplary application to the emotion shame and a schematic dynamic Bayesian network for its modeling are illustrated. Overall, this thesis provides evidence that human reactions towards Socially Interactive Agents are very similar to those towards humans, and that it is possible to model these reactions computationally.

Acknowledgments

This thesis could only have become what it is thanks to many people from different backgrounds who have contributed to this interdisciplinary work.

First of all, I would like to thank my supervisors, Dr. Patrick Gebhard and Prof. Dr. Cornelius König, for giving me the opportunity to write my dissertation on a topic that took some time to gain acceptance. Over the past few years, my relationship with Patrick has developed in such a way that we can easily be critical with each other, knowing that this relationship will not be disturbed. He has supported me along the way with invaluable empathy and enthusiastic encouragement. He is a fantastic boss, mentor and leader and I cannot think of a better leader for the Affective Computing Group. I sincerely thank Cornelius for his openness to the topics I have raised over the past few years and for his courage to go off the beaten track. His constant advice is irreplaceable and I am lucky to have him as a supervisor.

Further, I would like to express my gratitude to Prof. Dr. Birgit Lugrin for her willingness to serve as a second reviewer for this dissertation. Since I first met her, she has become a great role model for me as a woman doing research in computer science.

An office is where we get our inspiration and ideas, and where we spend much of our time. Therefore, I would like to pay special tribute to all my students, research assistants and colleagues who joined us in the cozy corridor on the first floor of E1 1. I would like to single out those truly exceptional people, Mirella, Bernhard, Anke, Manuel, Evelyn, Zhenqiang, Alvaro, Beka, Alexandra, Lydia, Sören, Naomi, Sofie, Lenny and Matthias, for making the time in and out of the famous hallway unforgettable. The new hallway at DFKI will be made famous by many new wonderful colleagues: Dimitra, Janet, Fabrizio, Lara, Nina, Ann-Kristin, Chirag and Mina. Thank you for being the best colleagues and dance partners I could ever imagine.

I would also like to thank my collaborators and co-authors, especially Dr. Tobias Baur and Alexander Heimerl in the HCAI laboratory at the University of Augsburg, headed by Prof. Dr. Elisabeth André. The papers and conferences where we presented our work would not have been possible without them. Elisabeth's superpower of knowing how to challenge us made us grow and improve with each of her questions.

Special thanks also go to Dr. Angela Castronovo (née Mahr), who lit the research fire in me and was the reason why I started working at DFKI.

Finally, but most importantly, I would like to thank my family for supporting me throughout my life. My mother, father and sister, who have made me who I am. My grandmother and grandfather, who made me curious. My husband, who stands by me in all of my crazy moments – and there are some. And, of course, our two daughters, who have made me learn more about affective reactions than I ever expected.

Contents

List of Figures

List of Tables

1
Description of author's contribution to joint work

1.1 Underlying Scientific Papers

Six published, peer-reviewed scientific publications form the basis of this dissertation. Under the supervision of Patrick Gebhard, I contributed as first author to five of these publications included in this dissertation. For one publication, I contributed as the second author. All publications are based on interdisciplinary work between psychologists and computer scientists. Therefore, they follow a process that includes conceptualization, operationalization, realization including its iteration cycles and validation in the form of a study. Due to the computer science focus, they are published at major conferences in the area of Affective Computing, User Interfaces, and Autonomous Agents that follow individual formatting rules (e.g., page limit). The following list provides an overview of the publications with their corresponding links and clarification of the authorship of each paper.

1. The first paper (Chapter 4) introduces the underlying interactive system PARLEY, which is based on previous work and has been implemented throughout this thesis. PARLEY aims to train difficult social situations in a safe environment with one or more Socially Interactive Agents and therefore combines research in the fields of psychology, computational emotion modeling, and social signal interpretation. It forms insofar the basis of this dissertation as it describes the underlying system and characteristics of the Socially Interactive Agents used for the studies presented afterwards. PARLEY was demonstrated at the 24th International Conference on Intelligent User Interfaces (IUI 2019) and published in the conference proceedings companion.

   Our main task, as Tobias Baur, Patrick Gebhard, and I, was to conceptualize the interaction design that transfers various aspects of natural human-human interactions to human-agent interaction. Additionally, published as well as unpublished studies were conducted under my supervision to test design parameters of the Socially Interactive Agent behavior, such as latency and frequency of mimicry behavior (Neumayr, 2018) or interruption handling (Gebhard et al., 2019c). Tobias Baur and Patrick Gebhard contributed to realizing the technical system. Tobias Baur and I were responsible for the paper's general writing tasks; all authors reviewed the paper. I demonstrated the system at the conference.

2. Chapter 5 presents the second contribution of this thesis, which is a study examining whether Socially Interactive Agents can elicit the social emotion of shame in a job interview. The results indicate that Socially Interactive Agents can elicit shame to the same amount as humans. This study provides insights into the impact of Socially Interactive Agents on humans and the possible role that agents can play. If a Socially Interactive Agent can elicit an emotion that usually depends on the presence of other humans, we can

assume that a Socially Interactive Agent can mentally represent a human object. This work was presented at the 8th International Conference on Affective Computing and Intelligent Interaction (ACII 2019) and published in the conference proceedings.

I conceptualized and operationalized the study with Patrick Gebhard and Markus Langer. Markus Langer contributed by developing the job interview situation, while Patrick Gebhard implemented the system. Mirella Scholtes, Bernhard Hilpert, and I were responsible for conducting the experiment. Mirella Scholtes and Bernhard Hilpert were involved in the observational coding of shame. The analysis and interpretation of the results were mainly my tasks. I contributed with the general writing task of the paper; all authors performed reviews of the paper. I presented the work at the conference.

3. The third paper (Chapter 6) presents an experiment on obedience that compares a Socially Interactive Agent and a human in the role of an instructor. Participants were asked to perform stressful and shameful tasks under the cover story of a creativity test. The results indicate that the Socially Interactive Agent has the same authority as the human instructor and is able to elicit the same level of negative feelings such as stress and shame. Likewise the study presented in Chapter 5, this study sheds light on the impact of Socially Interactive Agents on humans and the affective reactions towards them. Additionally, it replicates the result that Socially Interactive Agents can elicit the emotion of shame to the same amount as humans. This work was presented at the 8th International Conference on Affective Computing and Intelligent Interaction (ACII 2019) and published in the conference proceedings.

   Together with Patrick Gebhard and Sofie Ehrhardt, I conceptualized and operationalized the study. Sofie Ehrhardt developed and realized the stressful and shameful tasks. Patrick Gebhard and Manuel S. Anglet led the implementation of the system with support from Sofie Ehrhardt and me. Sofie Ehrhardt acted as experimenter. The analysis and interpretation of the results was mainly Sofie Ehrhardt's task with my support. I contributed with the general writing task of the paper; all authors performed reviews of the paper. I presented the work at the conference.

4. Chapter 7 contains the fourth contribution to this thesis. It presents a virtual biofeedback stress management training with a Socially Interactive Agent as a trainer and its evaluation. The goal was to develop a valid method for learning techniques on how to cope with stressful social situations in a safe environment with a Socially Interactive Agent. In an experiment, we compared our novel stress management training to a stress management training using stress diaries. The results indicate that our agent-based

stress management training using biofeedback significantly decreased the self-assessed stress levels immediately after the training, as well as in a socially stressful task with a human. Moreover, it seems to have positive effects on trainee's self rated performance in this task. This study shows that Socially Interactive Agents are not only able to elicit negative affects (cf. Chapter 5 & 6), but that it is also possible to practice strategies on how to cope with negative affect in difficult social situations with Socially Interactive Agents. This work was presented at the 26th International Conference on Intelligent User Interfaces (IUI 2021) and was honored as an outstanding paper. It was published in the conference proceedings.

With Naomi Sauerwein and Patrick Gebhard, I conceptualized and operationalized the study. Naomi Sauerwein developed the training content. Manuel S. Anglet and Patrick Gebhard supported by extending the PARLEY system with an interactive biofeedback environment that includes biosignal interpretation, a biofeedback monitor, and an image display. Naomi Sauerwein functioned as experimenter. The analysis and interpretation of the results was mainly Naomi Sauerwein's task with my support. I contributed with the general writing task of the paper; all authors performed reviews of the paper. I presented the work at the online conference.

A peer-reviewed position paper describing the development of the social biofeedback training system for stress management training was presented at the Workshop on Social Affective Multimodal Interaction for Health and published in the companion publication of the 2020 International Conference on Multimodal Interaction.

5. The fifth paper (Chapter 8) presents a computational model of user emotions for Socially Interactive Agents and its evaluation. The model combines a simulation of appraisal and emotion regulation processes for the emotion shame with social signal interpretation. The model's evaluation results show that social signal sequences can be related to emotion regulation processes. The corpus used to evaluate the computational model was gathered during the experiment, in which shame-eliciting situations were recorded (see Chapter 5). This paper shows how difficult user emotions without a clear emotional expression can be simulated and faced by Socially Interactive Agents. This work was presented at the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018) and published in the conference proceedings.

   With Patrick Gebhard and Tobias Baur, my main tasks were during the conceptualization of the computational model and its evaluation, including the operationalization and realization of the study. Additionally, it included the development of the annotation schema for the collected data. Patrick Gebhard's and Tobias Baur's task was to realize and evaluate the computational model. Patrick Gebhard, Tobias Baur and I contributed with the general writing task of the paper; all authors performed reviews of the paper.

6. The sixth paper (Chapter 9) presents the DEEP method as a starting point for a deeper computational modeling of emotions as internal, highly subjective experiences that are mostly not openly displayed. The method includes how to query individual internal emotional experiences, and it shows an approach to represent such information computationally. It combines social signals, verbalized introspection information, context information, and theory-driven knowledge. For an exemplary application on the emotion shame, a new corpus based on the shame-eliciting job interviews was recorded (see Chapter 5). In the end, a schematic dynamic Bayesian network for its modeling is illustrated.

   With Patrick Gebhard, I conceptualized and operationalized the data collection based on the job interview scenario in a former study (see Chapter 5). Mirella Hladký, Ann-Kristin Thurner and Jana Volkert supported with creating the interview for gathering the verbalized introspection information. I collected the data. The analysis and interpretation of the data were done by Mirella Hladký, Ann-Kristin Thurner and me. Alexander Heimerl, Tobias Baur and I conceptualized the dynamic Bayesian network. I contributed with the general writing task of the paper; all authors performed reviews of the paper. I presented the work during the online conference.

1.2 Paper Overview

The following table lists all peer-reviewed scientific publications that my colleagues and I have published until 2022 since I joined the Affective Computing Group. They are grouped by years, while the highlighted ones are the underlying scientific publications of this thesis.

Table 1.1: Published peer-reviewed papers.

2022	Gebhard, P., Tsovaltzi, D., Schneeberger, T., & Nunnari, F. (2022). Serious games with SIAs. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 2: Interactivity, platforms, application* (pp. 527–546). https://doi.org/10.1145/3563659.3563676

Wessler, J., Schneeberger, T., Christidis, L., & Gebhard, P. (2022). Virtual backlash: Nonverbal expression of dominance leads to less liking of dominant female versus male agents. *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 1–8. https://doi.org/10.1145/3514197.3549682 ***Best Paper Award***

Heimerl, A., Mertes, S., Schneeberger, T., Baur, T., Liu, A., Becker, L., Rohleder, N., Gebhard, G., & André, E. (2022). Generating personalized behavioral feedback for a virtual job interview training system through adversarial learning. *Proceedings of the 23rd International Conference on Artificial Intelligence in Education*, 679–684. https://doi.org/10.1007/978-3-031-11644-5_67

2021 Schneeberger, T., Aly, F. A., Don, D. W., Gies, K., Zeimer, Z., Nunnari, F., & Gebhard, P. (2021). Influence of movement energy and affect priming on the perception of virtual characters extroversion and mood. *Companion Publication of the 23rd International Conference on Multimodal Interaction*, 211–219. https://doi.org/10.1145/3461615.3485409

Wessler, J., Schneeberger, T., Hilpert, B., Alles, A., & Gebhard, P. (2021). Empirical research in affective computing: An analysis of research practices and recommendations. *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction*, 1–8. https://doi.org/10.1109/ACII52823.2021.9597418

Schneeberger, T., Hladký, M., Thurner, A.-K., Volkert, J., Heimerl, A., Baur, T., André, E., & Gebhard, P. (2021). Towards a deeper modeling of emotions: The DEEP method and its application on shame. *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction*, 1–8. https://doi.org/10.1109/ACII52823.2021.9597446

Hladký, M., Schneeberger, T., & Gebhard, P. (2021). Understanding shame signals: Functions of smile and laughter in the context of shame. *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction: Workshops and Demos*, 1–7. https://doi.org/10.1109/ACIIW52867.2021.9666424

	Schneeberger, T., Sauerwein, N., Anglet, M. S., & Gebhard, P. (2021). Stress management training using biofeedback guided by social agents. *Proceedings of the 26th International Conference on Intelligent User Interfaces*, 564–574. https://doi.org/10.1145/3397481.3450683 ***Honorable Mention***
	Muscholl, N., Klusch, M., Gebhard, P., & Schneeberger, T. (2021). EMIDAS: Explainable social interaction-based pedestrian intention detection across street. *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 107–115. https://doi.org/10.1145/3412841.3441891
2020	Schneeberger, T., Sauerwein, N., Anglet, M. S., & Gebhard, P. (2020). Developing a social biofeedback training system for stress management training. *Companion Publication of the 22nd International Conference on Multimodal Interaction*, 472–476. https://doi.org/10.1145/3395035.3425222
	Kruijff-Korbayova, I., Hackbarth, J., Jacob, C., Kiefer, B., Schmitt, M., Schneeberger, T., Schwartz, T., Horn, H.-P., & Bohlmann, K. (2020). *Towards intuitive verbal and non-verbal communication for incidental robot-human encounters in clinic hallways.* Retrieved December 9, 2022, from https://www.dfki.de/fileadmin/user_upload/import/11125_IKK_etal_Towards_Intuitive_Verbal_and_Non-Verbal_Communication_for_Incidental_Robot-Human_Encounters_in_Clinic_Hallways.pdf
2019	Schneeberger, T., Scholtes, M., Hilpert, B., Langer, M., & Gebhard, P. (2019). Can social agents elicit shame as humans do? *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction*, 164–170. https://doi.org/10.1109/ACII.2019.8925481
	Schneeberger, T., Ehrhardt, S., Anglet, M. S., & Gebhard, P. (2019). Would you follow my instructions if I was not human? Examining obedience towards virtual agents. *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction*, 1–7. https://doi.org/10.1109/ACII.2019.8925501
	Schneeberger, T., Gebhard, P., Baur, T., & André, E. (2019). Parley: A transparent virtual social agent training interface. *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, 35–36. https://doi.org/10.1145/3308557.3308674

Schneeberger, T., Hirsch, A., König, C., & Gebhard, P. (2019). Impact of virtual environment design on the assessment of virtual agents. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 148–150. https://doi.org/10.1145/3308532.3329455

Gebhard, P., Schneeberger, T., Mehlmann, G., Baur, T., & André, E. (2019). Designing the impression of social agents' real-time interruption handling. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 19–21. https://doi.org/10.1145/3308532.3329435

Gebhard, P., Schneeberger, T., Dietz, M., André, E., & Bajwa, N. u. H. (2019). Designing a mobile social and vocational reintegration assistant for burn-out outpatient treatment. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 13–15. https://doi.org/10.1145/3308532.3329460

2018 Schneeberger, T. (2018). Transfer of social human-human interaction to social human-agent interaction. *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, 1778–1780.

Gebhard, P., Schneeberger, T., Baur, T., & André, E. (2018). MARSSI: Model of appraisal, regulation, and social signal interpretation. *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, 497–506.

Gebhard, P., Schneeberger, T., André, E., Baur, T., Damian, I., Mehlmann, G., König, C., & Langer, M. (2019). Serious games for training social skills in job interviews. *IEEE Transactions on Games*, 11(4), 340–351. https://doi.org/10.1109/TG.2018.2808525

2
General Introduction

Our world is a social place, and humans are social beings who need social bonds. The need to belong is a fundamental characteristic of all human beings. It motivates human actions and leads to regular social interactions between them (Baumeister & Leary, 1995). The human way of interaction relies on one's own emotions, and those of others. Experienced and expressed emotions help us to built and maintain social relationships (Fischer & Manstead, 2008) as they inform ourselves and the counterpart about intrapersonal processes and mechanisms (Frijda, 2008). It is difficult to imagine any single interaction that does not include communicating some affective information, receiving affective information, and/or experiencing some affective state (Halberstadt et al., 2001). However, for successful interactions and social relationships, a mutual understanding of interaction partners is required (Moser, 2013).

Next to human-human interaction, the fast development of personal computers in the last three decades has led to an increased amount of human-computer interaction. While research in human-computer interaction was first driven by a strong emphasis on tasks and usability, it got later enriched by taking affective reactions of users towards the computer into account (Diefenbach et al., 2014). One approach to creating natural and intuitive interactions between humans and computers is to develop artificial agents capable of interacting through communication channels similar to those of humans. These so-called Socially Interactive Agents (SIAs) act as the machine interface enabling a multi-modal human-computer interaction by using verbal, para-verbal, and non-verbal behaviors. This makes it possible to transfer communication styles that are known from face-to-face human interaction to human-computer interaction (Lugrin, 2021).

Whether the reactions towards computers are similar to those towards humans is studied within the Computers-Are-Social-Actors paradigm (Nass et al., 1997; Nass et al., 1994). Studies following this approach compare social reactions shown towards a computer with reactions towards humans. Previous work has found that humans show communication behavior in human-SIA interactions that is equivalent to that expected in face-to-face conversations (Gratch et al., 2007; Hoffmann et al., 2009; Kopp et al., 2005; Krämer et al., 2013). Moreover, affective reactions towards SIAs show that they are perceived as social entities (Bickmore et al., 2020; Lucas et al., 2014; Pauw et al., 2022; Weitz et al., 2019, 2021).

Although there are already studies examining human reactions towards SIAs, it remains an open research question to what extent affective reactions towards them resemble the ones towards humans. However, this is a crucial question as it defines future applications of SIAs and the possible roles they can overtake in humans' lives. Also, it is still unclear how to computationally model these affective reactions. However, during human-SIA interactions this knowledge is important as it lays the basis for social interactions. As a beginning to approach these questions, this work examines in three studies, firstly, if a SIA can elicit affective reactions that are so far known to be only elicited by a human (Chapter 5). Secondly, if humans obey the instructions of SIAs the way they obey instructions of another human (Chapter 6) and finally, if SIAs can be employed to teach techniques on how to cope with stressful social situations (Chapter 7). Additionally, it presents the general system with the SIAs used in the previously mentioned studies (Chapter 4). In the end, two approaches how to model user affect computationally are presented (Chapter 8 & 9).

3
General Theoretical Background

In order to understand what is meant by Socially Interactive Agents (SIAs), the purpose of their application and the terminology are described below. Furthermore, this chapter presents studies that investigate human reactions towards SIAs and systems that use SIAs for training purposes. Finally, theories explaining human social behavior towards SIAs are described, leading to the open research questions of this dissertation thesis.

3.1 Socially Interactive Agents

The initial goal of applying virtual agents in the field of human-computer interaction was to enable human users to interact via communication channels that come naturally to them (Cassell et al., 2000). By adding humanoid aspects and equipping the interface with a body that interacts multi-modally, virtual agents communicate with users verbally, para-verbally, and non-verbally (Pelachaud, 2009). This development incorporates human-like cues into the interface, and with this new social dimensions enter human-computer interaction (Krämer et al., 2018). The transfer from communication styles that are known from human face-to-face interaction to the interaction with machines results in a more human-like interface that is intuitive to understand and to interact with. In recent years, the use of virtual agents has been extended beyond the improvement of human-computer interaction. Applications supporting users by exploiting virtual agents as companions or assistants were developed. Virtual agents are simulating experts, such as interaction partners or advisors in the health context (Bickmore et al., 2010; DeVault et al., 2014; Gebhard et al., 2019b), in museums (Bickmore et al., 2011; Bickmore et al., 2013b) or for vulnerable groups (Bickmore et al., 2005a; Burke et al., 2018; Milne et al., 2010) to provide on-demand support. Moreover, virtual agents are exploited in social training systems to simulate human-like dialog partners for the trainee to practice difficult social situations. In perception studies, virtual agents can function as highly standardized stimulus material. In interaction studies, virtual agents can serve as confederates as they ensure to express the same behavior over the course of the study. Therefore, virtual agents are used to studying aspects of human-human interaction in social psychology (Bailenson & Yee, 2005; Pan & Hamilton, 2018), the learning sciences (Bailenson et al., 2008), or the language sciences (Peeters, 2019).

There is no consistent term or definition of what a SIA is. They have been developed under different names in different research fields, whereas each field and term comes with its own prioritization of aspects. Examples are Intelligent Virtual Agents (Rickel, 2001), Embodied Conversational Agents (Cassell et al., 2000), Socially Adaptive Virtual Agents (Youssef et al., 2015), Socially Intelligent Agents (Bickmore, 2003; Dautenhahn, 1998a, 1998b), or Virtual Humans (DeVault et al., 2014; Traum et al., 2008). In general, an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators (S. Russell & Norvig, 2002). Hence, agents can be, but are not necessarily, embodied and are characterized by a high degree of autonomy. They are fully controlled by algorithms (Pan & Hamilton, 2018). On the other hand, an avatar represents a character that is fully controlled in real-time by another person (Bainbridge, 2004; Pan & Hamilton, 2018). It often represents a game unit that is under the player's control (Kromand, 2007), which is usually the graphical representation of the user in the virtual environment (Trepte & Reinecke, 2010). An avatar does not behave or interact autonomously with a user but represents a user in the virtual world. It is used in games and experiments (Lucas et al., 2014), or in the computer-mediated workplace conversation (Inkpen & Sedlins, 2011).

Pan and Hamilton (2018) define a two-dimensional space to classify existing agents on the axes graphical realism (x-axis) and interaction dynamics (y-axis). Graphical realism ranges from simple pixelated characters (low graphical realism) to photorealistic characters (high graphical realism). Interaction dynamics is described as the level of interactivity between the participant and the computer system and ranges from non-responsive to fully responsive. The more pronounced both dimensions are, the more an agent resembles a real human. At one extreme, the authors classify the movie Avatar (Cameron, 2009). The characters are photorealistic, but there is no possibility of interacting with them as the story progresses. At the other extreme, the authors classify the computer game Pacman. The very simple pixelated characters are highly responsive to both keystrokes and each other.

Based on the classification of Pan and Hamilton (2018), the agents exploited in the studies presented in this thesis are quasi-agents. These are characters that are partly autonomous and partly controlled by a human. However, by applying the Wizard-of-Oz technique (Kelley, 1984), participants get the impression of interacting with an autonomous agent, as an experimenter (Wizard) controls several aspects of the virtual agent without the participants' knowledge. Moreover, the virtual agents used for the following studies can be understood as SIAs. They have a virtual representation of a figure along with animations that resemble a human. They can interact with humans using social communicative behaviors common to human-human interaction, using both verbal and non-verbal signals. Not only do they behave socially, but they can also recognize and identify other agents and establish and maintain relationships with other agents, whereas these other agents are not necessarily computational agents (Dautenhahn, 1998a), but in our case, the human user. SIAs can perceive verbal and nonverbal cues and subsequently react to the given input. They are equipped with feedback and turn-taking features. Moreover, they can engage the user in a relevant conversation using social cues such as speech, gestures, gaze, and facial expressions (see Chapter 4 for the technical description of the Socially Interactive Agent system).

3.2 Social Behavior Towards Socially Interactive Agents

The Computers-Are-Social-Actors paradigm (Nass et al., 1997; Nass et al., 1994), derived from the media equation (Reeves & Nass, 1996), suggests that humans treat media and computers like real people, mindlessly applying scripts for interacting with humans to interactions with social technologies. This research has documented that people's responses to computers are fundamentally "social" – that is, people apply social rules, norms, and expectations core to interpersonal relationships when interacting with computers. Nass and Reeves themselves conducted several empirical studies applying the CASA paradigm. They found several examples of social behavior towards computers: A computer is getting evaluated better if it is praised by another computer than by itself (Nass & Steuer, 2006). People would rather help a computer that has helped them before than a computer that has not helped them before (Nass & Moon, 2000). Also, gender stereotypes are transferred to computers by finding that a computer with a female voice knows more about love and relationships than a computer with a male voice (Nass & Moon, 2000).

To trigger social reactions, the computer, which the user interacts with, needs to elicit certain social cues. The stimuli considered especially effective in evoking a human association have been grouped into three categories: speech as output modality, interactivity (i.e., responses based on previous interactions), and the filling of roles traditionally filled by humans (Nass & Moon, 2000). Therefore, it is not surprising that adding a human-like virtual agent to the computer interface results in an even more pronounced social behavior towards computers (Krämer et al., 2018).

Numerous studies yield social effects, demonstrating that humans' reactions towards virtual agents are remarkably similar to those towards human interlocutors. In the following, several results from both authors with psychology, as well as computer science background, examining social reactions towards human-like virtual agents will be presented to provide an overview of the research.

3.2.1 Behavioral Reactions Towards Socially Interactive Agents

SIAs evoke communication behavior in humans that is equivalent to that expected in a face-to-face conversation. This includes human-like communication strategies, cooperative behavior (Kopp et al., 2005), polite behavior (Hoffmann et al., 2009), and rapport building (Gratch et al., 2007).

In a field study, Kopp et al. (2005) used a conversational agent as a guide in a public museum. The agent engaged with visitors in natural face-to-face communication, provided them with information about the museum or the exhibition, and conducted natural small talk conversations. The analysis of conversations showed that museum visitors accepted the agent as a conversation partner. Visitors showed human-like communication strategies (e.g., greeting) and behaved cooperatively towards the agent. Moreover, visitors asked many anthropomorphic questions and tried to flirt with the agent. The authors concluded that the visitors' engagement indicates the attribution of sociality to the agent (Kopp et al., 2005). Using the same virtual agent, Hoffmann et al. (2009) examined how participants evaluated the agents after a 10-minute conversation. The evaluation was either done by being questioned by the agent himself, being questioned by paper-and-pencil questionnaire in the same room facing the agent, and being questioned by means of a paper-and-pencil questionnaire in another room. When the agent was interactively asking participants how they would evaluate him, participants gave a better evaluation which the authors describe as more polite behavior. Participants seemed to have difficulty giving negative feedback face-to-face to the assessee (Hoffmann et al., 2009).

Gratch et al. (2007) found evidence that an agent engenders feelings of rapport in human speakers. They compared a human to an agent designed to elicit rapport within a dyadic narrative task. The rapport agent provided nonverbal listening feedback associated with rapportful interactions. Those included backchannelling behavior (e.g., nods), postural mirroring, and mimicry of certain head gestures (e.g., gaze shifts and head nods). Their study results indicate that the rapport agent was as good as human listeners in creating rapport (Gratch et al., 2007). However, in 2018, this result could not be replicated (Krämer et al., 2018).

Krämer et al. (2013) let participants small talk for eight minutes with a virtual agent. The virtual agent either did not smile, showed occasional smiles, or displayed frequent smiles. The study results showed that though the smiling behavior of the agent has not been perceived consciously, it influenced participants smiling behavior. When the agent was smiling, the duration of the participants' smiling was longer. Also, smiling behavior did not affect the evaluation of the agent. The authors concluded that participants' behavioral reactions were rather unconscious and automatic (Krämer et al., 2013). Though smiling has been analyzed within the mimicry paradigm (Chartrand & Bargh, 1999; Chartrand et al., 2005), the authors stress that this reciprocation might not be defined as mimicry as other mechanisms can come into play (e.g., politeness rules, usage as a communication facilitator). However, smiling behavior has a special function in interpersonal interactions. Expressing emotions, a smile can represent a major component of a facial display that might be associated with and caused by feelings of happiness or joy (Ekman & Friesen, 1971). Moreover, which might be even more crucial, it regulates the relationship between interaction partners. Smiling has strong and robust associations with social motivation and is an important mean of communication (Kraut & Johnston, 1979). Therefore, it seems that participants experienced the small talk with a virtual agent as a social situation.

Overall, it can be summarized that numerous studies find evidence of a similar human communication behavior towards SIAs and humans. Therefore, one could assume that in future studies with SIAs as interactants, researchers will also observe communication behavior that is similar to human-human interactions.

3.2.2 Affective Reactions Towards Socially Interactive Agents

SIAs activate not only human-like conversational behavior in human interaction partners, but also similar affective responses like in human-human interaction.

Comparing a text interface with a conversational interface having a face, Sproull et al. (1996) found evidence that a more human-like agent can affect the affective state and behavior of a human interaction partner. In an interaction with a computer career counseling service, participants attributed personality to the conversational interface differently than to the text display. Moreover, they reported higher arousal and presented themselves more positively when interacting with the conversational interface (Sproull et al., 1996).

Deladisma et al. (2007) observed medical students conducting a medical interview with an agent in the role of a patient with pain. To trigger empathic responses, the virtual patient expressed his fear and asked for help. Though less than with a standardized human patient, the students demonstrated nonverbal communication behaviors and responded empathetically to the virtual patient. This empathic response indicates that the students appreciate the agent's emotional situation and try to build a common understanding of the illness (Deladisma et al., 2007).

The presence of a human-like virtual agent also seems to have effects on users' trust (Weitz et al., 2019, 2021). In two experiments, the authors studied whether explainable artificial intelligence visualizations profited from the enrichment with virtual agents. The results of the earlier study showed that integrating a human-like virtual agent that explains complex facts led to increased trust in an autonomous intelligent system (Weitz et al., 2019). In the later study, the authors found evidence that the more human-like explainable artificial intelligence interactions appeared, the more the users tended to trust the classification model whose predictions were explained (Weitz et al., 2021). The virtual agent used for this experiments is the same one used in the studies presented in Chapters 6 and 7 of this thesis.

One reason for the high trust of humans in SIAs might be that they are experienced as supportive and safe interaction partners. Lucas et al. (2014) studied the potential of SIAs as interviewers. In their study, participants were led to believe that a SIA conducting a semi-structured health screening was controlled by either a human or automation. Their results showed that the automated agent evoked lower fear of self-disclosure and lower impression management. Moreover, participants interacting with an agent displayed negative emotions more intensely and were rated by observers as more willing to disclose (Lucas et al., 2014). In a follow-up study, they showed that service members after a year-long deployment in Afghanistan reported more openly to a SIA about posttraumatic stress disorder symptoms compared to a questionnaire (Lucas et al., 2017). In both studies, the SIAs were designed to build rapport. Human-human rapport is a subjective experience of attunement between interactants that is strongly connected to nonverbal behavior (Tickle-Degnen & Rosenthal, 1990). Its various positive influences on the interpersonal process have driven researchers to recreate this interpersonal state within human-agent interactions, especially by creating agents that show appropriate backchanneling behavior (DeVault et al., 2014; Gratch et al., 2013; Gratch et al., 2006; Gratch et al., 2007).

Within a similar use-case of disclosing stigmatized information, Bickmore et al. (2020) developed and validated a virtual agent designed to automate the administration of a substance use screening instrument. In two studies, they found that the agent led to more disclosure compared to a human interviewer and to more satisfaction compared to a text-based tool. The qualitative data revealed that the agents' superiority lay in their non-judgmental way of conducting the screening (Bickmore et al., 2020). The finding replicates the results of a former study that revealed the preference to disclose negative, personally sensitive information to a virtual agent is mainly driven by their lack of judgment, criticism, as well as verbal or nonverbal reactions (Pickard et al., 2016).

Krämer et al. (2018) examined whether interactions with a virtual agent are experienced as socially rewarding and can meet social needs like human-human interactions. In their experiment, a virtual agent that asked the participant five questions with increasing intimacy either displayed socially responsive nonverbal behavior or not. Their results showed an effect of individual differences in the need to belong. For participants with a high need to belong, the interaction with a virtual agent lowered their intention to engage in social contact, but only for the agent that displayed socially responsive behavior.

It seems that conversations with virtual agents are not only socially rewarding but also socio-emotionally supporting (Pauw et al., 2022). Study results showed that after talking about two negative emotions, anger and worry, and getting emotional and cognitive support, participants felt better – the target emotion was reduced, and the affect was generally improved. This led the authors to the conclusion that talking to a virtual human can be a valuable form of support at times of distress.

The sum of research on interactions between humans and SIAs has shown that they are perceived as social entities. SIAs do not only activate human-like conversational behavior in human interaction partners, but also affective responses that would be expected in human-human interactions. Studies in the context of disclosing relevant information in healthcare indicate that SIAs can help to overcome a critical barrier, as they guarantee anonymity while also building rapport. As they also seem to give socio-emotional support, they can be applied to gathers sensitive mental health information of patients.

3.3 Applications of Socially Interactive Agents – Social Training Systems

Social training systems make use of human social behavior towards SIAs. In social training systems, SIAs represent social interaction partners and enable users to experience difficult social situations virtually. Belonging to the class of serious games, social training systems have seen rapid evolution in recent years due to advances in the areas of social signal processing as well as improvements in the audio-visual rendering of virtual agents (Gebhard et al., 2019a). Social training systems are realized either in immersive or non-immersive virtual reality (VR), a non-invasive simulation technology that allows users to interact in real-time with a computer-generated environment (Burdea & Coiffet, 2003). By employing role play with virtual agents, the goal of a social training system is to foster reflection about socio-emotional interactions and to complement or even substitute traditional training approaches (Gebhard et al., 2019a).

Social training systems, as a form of computer-based training, have several advantages. First, a social training system always provides the possibility of training as there is no actual human needed, such as a role player, which makes applying them very economical (Schmid Mast et al., 2018). Secondly, the trainee can practice difficult social situations in a protected environment, without human trainers or interlocutors (Gebhard et al., 2019a), which might reduce negative experiences (cf. 3.2.2). Thirdly, social training systems provide incomparable systematic and structured learning (Bedwell & Salas, 2010), as even with highly trained role players, it is challenging to have the same amount of control. Lastly, the scaffolding, flexibility, and adaptivity of social training systems, can enhance the transfer to real situations (Schmid Mast et al., 2018).

A variety of social training systems employing role play with virtual agents foster reflection about socio-emotional interactions. In the following, the focus will be on social training systems realized in non-immersive virtual realities. In those, the virtual reality is presented on a screen, including the SIAs with which the user can interact in real-time. Consequently, training systems that are realized in high-immersive virtual realities [e.g., Hartholt et al. (2019), Schmid Mast et al. (2018), Stansfield et al. (2000), and Thordarson and Vilhjälmsson (2019)] as well as interactive learning environments [for an overview, see Johnson et al. (2000)] are omitted here.

In several social training systems, virtual agents are simulating interviewers in the context of job interview training (Anderson et al., 2013; Damian et al., 2015a; Hoque, 2012; Hoque et al., 2013; Langer et al., 2016; M. J. Smith et al., 2014). The goal is to enable users to interact with a virtual job interviewer while getting feedback on their paraverbal and nonverbal behaviors. Studies show that training with such systems is subjectively and objectively beneficial. In an evaluation of the *EmpaT* job interview trainer, Langer et al. (2016) found that participants self-reported less interview anxiety and outperformed participants receiving conventional training in a job interview with a real interviewer. An evaluation of the *MACH* social skills training showed that providing visual feedback on the user's performance after a job interview with the virtual coach improved general interview performance (Hoque et al., 2013). Overall, there is evidence that training systems combining virtual agents with automatic feedback on paraverbal and nonverbal behavior can successfully help with job interview preparation.

The anti-bullying game *FearNot!* aims to enhance social learning through interactive role play with virtual agents that establish empathetic relationships with the learners (Hall et al., 2009). It creates interactive stories in a virtual school with embodied conversational agents, for example in the role of bullies, helpers, or victims. The children run through various bullying episodes, interact with the virtual agents after each episode, and provide advice to them. The benefit of educational role plays of this kind lies in the fact that they promote reflective thinking. Results of a conducted evaluation (Sapouna et al., 2010) showed that the system positively affected the children's abilities to cope with bullying.

Another difficult social situation that can be trained with social training systems is public speaking (Batrinca et al., 2013; Chollet et al., 2014; Chollet et al., 2015; Slater et al., 1999). Slater et al. (1999) wanted mainly to explore the effectiveness of virtual environments in psychotherapy for social phobias. Instead of realizing and evaluating a virtual reality therapy tool, they developed a public speaking trainer, as public speaking is a prevalent cause of anxiety among the general population. In their experiment, they examined whether the interest intensity of a virtual audience influenced participants' self-rated performance and public speaking anxiety. The audience's interest intensity was manipulated by their behavior, such as showing happy or unhappy faces, applauding, or talking to each other. The authors found that higher perceived audience interest increased self-rating and reduced public speaking anxiety. Another system for practicing presentations in a safe and engaging environment is the *Cicero* platform (Chollet et al., 2014). The researchers' aim was to develop a system that is able to give automatic feedback to the user about her performance. Therefore, they identified several characteristic nonverbal behaviors that correlated positively or negatively with the overall expert-rated presenters' performance (Batrinca et al., 2013). Moreover, they investigated the optimal ways of conveying the perceived information on the performance to the presenters (Chollet et al., 2015). One of the evaluation studies of *Cicero*, a pre- vs. post-training evaluation, confirmed its positive influence on three assessment perspectives: (1) the presenters themselves, (2) public speaking experts, and (3) objectively annotated behavioral data. Overall, the authors concluded that the training with *Cicero* improved trainees' public speaking by developing good nonverbal communication skills. Moreover, a repeated confrontation with the situation may have regulated or have reduced public speaking anxiety. In the main, training systems for public speaking focus on rehearsing beneficial nonverbal behaviors, like flow of speech or eye gazing. Practicing beneficial nonverbal behavior might improve not only the externally rated performance in public speaking but also the trainees' subjective assessment, such as self-rated performance. However, this has yet to be explicitly the goal of public speaking training systems.

Ali et al. (2018, 2020) employed social agents as conversational partners for groups of people with difficulties in social communication. The system *Aging and Engaging* used a virtual agent to help improve social skills among older adults (Ali et al., 2018). Users were talking with an agent that looks like an older adult about family and friends. After interacting with the agent, users received automated feedback using a hidden Markov model on eye contact, volume, smile, and speech content. The researchers followed a similar approach when developing a virtual conversational agent for teens with autism spectrum disorder (Ali et al., 2020). Within the training, users conversed informally with a virtual agent, receiving feedback on nonverbal cues in real-time and summary feedback. In a pilot study, the usefulness of the feedback and the dialogues were assessed positively. Several other interventions have proven the applicability of interactive technologies in social skills training for individuals with autism spectrum disorder (Bernardini et al., 2013; Foster et al., 2010; Hopkins et al., 2011; Mower et al., 2011; Razavi et al., 2016; Tartaro & Cassell, 2008). Similar to public speaking training, these systems focus on changing nonverbal behaviors that are measured automatically by the system and not on users' subjective experience in the situation.

Several virtual agents embodying the role of virtual patients have been proposed to train medical students or novice clinicians. Within these training systems, users practice the assessment of patients with mental disorders (Kenny et al., 2008), communication skills (Johnsen et al., 2006; Lok et al., 2006; Stevens et al., 2006), or how to deliver bad news (Andrade et al., 2010; Ochs et al., 2019). Overall, it seems that virtual patients can successfully be exploited to train unexperienced clinicians, also because doctors demonstrate non-verbal behaviors and respond empathetically to virtual patients (Deladisma et al., 2007). Moreover, these social training systems provide a controllable, secure, and safe learning environment with the opportunity for repetitive practice.

One more use case for social training systems is intercultural competence. Social training systems with SIAs can enhance acceptance of foreign cultures, raise cultural awareness, teach about cultures and culture-specific differences, foster cross-cultural understanding and reduce subconscious biases (Lugrin & Rehm, 2021). Both, in US (Deaton et al., 2005; Johnson & Valente, 2008; Kim et al., 2009; Lane et al., 2013; Lane et al., 2008; Raybourn et al., 2005), and Europe (Aylett et al., 2009; Endrass et al., 2013; Hall et al., 2015; Mascarenhas et al., 2013; Nazir et al., 2012), research groups have developed several applications to train intercultural communication using virtual agents. The systems developed in US mostly focus on training cultural competencies to military members for operations abroad, whereas the European training systems focus on training civil children, adolecents and adults in general cultural understanding. Likewise for other social training systems, the advantages of employing SIAs are that the task can be repeated as often as necessary while keeping an emotional distance (Lugrin & Rehm, 2021).

In sum, this body of work suggests that users can successfully train difficult social situations with social training systems. Most systems focus on the training of a specific paraverbal and nonverbal behavior. The training systems offer the potential for users to explore social situations and practice different behavior responses for a variety of simulated social interactions (Kerr et al., 2002).

3.4 Explaination for Social and Affective Reactions Towards Socially Interactive Agents

As outlined before, humans express social and affective reactions towards SIAs similar to those shown towards humans. This is an essential prerequisite for the successful application of SIAs in social training systems. Different models try to explain the roots of this social and affective behavior (Krämer & Manzeschke, 2021).

The most prominent explanation for social behavior towards computers is the "ethopoeia" approach in the media equation assumption (Nass & Moon, 2000; Reeves & Nass, 1996). Ethopoeia involves a direct, mindless response to an entity as human while knowing that the entity does not warrant human treatment or attribution. Users automatically and unconsciously apply social rules to their interactions with computers - since humans are inherently social and computers display social cues (Nass et al., 1997). The social reactions are shown even though they conflict with the informed users' opinion - computers do not require social treatment. Social reactions are triggered according to this approach already by minimal social cues like human-sounding voice. The media equation assumption is backed by the authors' numerous studies following the CASA paradigm: replicate an experiment from social science and observe whether social rules are still applied when "human" is replaced by "computer" in the experiment. In the experiments, indeed, participants did, firstly, not consciously recognize their social behavior towards the computer. Secondly, when asked after the experiment, participants declared not having acted socially and found such behavior unsuitable. The authors concluded that their findings are verifying the media equation: users perceive technologies as social actors as they transfer human-human reaction scripts to human-computer interaction (Nass et al., 1994).

Over the years, research tried to find evidence that the reactions towards computers are not truly social. First, it was hypothesized that social reactions were due to psychological dysfunctions, young age, or lack of computer experience. Then as well it was argued that users thought they were rather interacting with some projected human “behind” the computer. Another explanation was that social behavior appears rather due to missing scripts during the human-computer interaction (Kiesler & Sproull, 1997). However, these assumptions could be rejected by several studies. When applying social rules, it seems that individuals directly interact with the computer as an independent social actor or source and not with some projected human “behind” the computer, such as the programmer (Nass et al., 1999; Sundar & Nass, 2000). The computer is not perceived as a channel for communication between the user and other humans, where the polite responses were entirely appropriate (Nass et al., 1999). Moreover, participants are unconscious about their social behavior towards agents (Krämer et al., 2013).

Finally, it can be concluded that the psychologically relevant source in human-computer interaction can be only the computer itself to which the human user reacts automatically and unconsciously social. This becomes especially obvious due to the contradiction between the actual display of social reactions and the explicit conviction that such reactions are irrational (Nass & Moon, 2000) – supporting the assumptions made based on the Computers-Are-Social-Actors paradigm (Nass et al., 1997; Nass et al., 1994).

3.5 Computational Emotion Models for Socially Interactive Agents

To create believable behavior for SIAs in social training systems, computational models simulate affective reactions of human interaction partners. Computational emotion models are mathematical or algorithmic representations of human emotions used to understand and simulate emotions (Ojha et al., 2021). Originating in the field of psychology, few cognitive appraisal theories for emotions were the starting point for modeling emotions computationally in the area of Affective Computing (Moors et al., 2013).

One of the earliest and most influential emotion models is the Circumplex Model of Affect, proposed by James Russell in 1980. This model represents emotions as points in a two-dimensional space defined by valence (pleasantness-unpleasantness) and arousal (activation-deactivation). Valence refers to the positivity or negativity of an emotional experience, while arousal refers to the level of physiological activation or stimulation associated with the experience (J. A. Russell, 1980). Later, various models have been proposed that build on this foundation, such as the PAD (Pleasure-Arousal-Dominance) model (Mehrabian, 1996) or the Geneva Emotion Wheel (Bänziger et al., 2005). The PAD model describes the emotional state in terms of three basic dimensions: pleasure, arousal, and dominance. According to the model, pleasure represents the degree to which an individual experiences positive or negative feelings, arousal represents the degree of physiological activation or stimulation, and dominance represents the degree of control or power the individual feels in a situation. Similarly, the Geneva Emotion Wheel is based on three dimensions: valence, arousal, and control. It adds emotion terms in a systematic fashion by aligning them with respect to the underlying dimensional structure (Bänziger et al., 2005).

Another model, originally from psychology, is the OCC model from Ortony, Clore, and Collins (1988). It proposes that emotions are cognitive in nature and are based on an individual's appraisal of the current state of the world. In other words, the model suggests that emotions arise from the way individuals perceive and interpret events in their environment. The OCC model has three components: 1) object appraisal: an individual's evaluation of an object or situation in terms of its relevance to their goals, values, and needs; 2) consequence appraisal: an individual's evaluation of the likely consequences of the object or situation in terms of their goals, values, and needs; 3) appraisal of coping potential: an individual's evaluation of their ability to cope with the object or situation in terms of their goals, values, and needs (Ortony et al., 1988).

The computational modeling of emotions started in the 1980s (Pfeifer, 1988). Computational models of human emotional processes enable SIAs to exhibit emotions and interact with humans empathically by simulating and responding to human emotions appropriately (Marsella et al., 2010; Vinayagamoorthy et al., 2006). The most prominent emotion models for SIAs based on the aforementioned emotion models are EMA (Gratch & Marsella, 2005), FAtiMA (Dias et al., 2014), and ALMA (Gebhard, 2005).

EMA is used by empathic agents in various systems (e.g., Swartout et al. 2006) to model appraisal and reappraisal of users (Marsella & Gratch, 2014). The model aims to account for both the factors that give rise to emotions as well as the wide-ranging impact emotions have on cognitive and behavioral responses, particularly coping responses (Gratch & Marsella, 2005). Although EMA is very powerful, it is not able to model complex emotions, like social emotions, such as pride and shame (Lewis, 2008).

FAtiMA is a generic and flexible architecture for SIAs. It includes a set of appraisal rules that determine the emotional significance of stimuli and a set of decision rules that determine the behavioral responses of the virtual agent (Dias et al., 2014).

ALMA integrates three major affective characteristics: emotions, moods, and personality, that cover short, medium, and long-term affect. It provides SIAs with a personality profile and with real-time emotions and moods to generate their multimodal behavior (Gebhard, 2005).

Overall, most of the current computational models of emotions follow the concept of cognitive appraisal-based emotion elicitation. Some models also include coping. However, the explicit modeling of complex social emotions and the relation of observed social signals to situational appraisal regulation representations still remain an open research challenge.

3.6 General Aim of this Dissertation

While the framework of the Computers-Are-Social-Actors paradigm has been widely applied to examine users' social reactions to technologies, the mechanisms explaining users' social reactions are still unclear (Xu et al., 2022). Moreover, due to the fast development of robots, there has been much research on robots as interaction partners (Appel et al., 2021; Cheok & Levy, 2018; Cooper et al., 2020; Dautenhahn et al., 2005; Gambino et al., 2020; Moshkina et al., 2014), but less with SIAs as interfaces. Therefore, the work done for this dissertation aims to explore more facets of how humans socially behave towards machines that have a SIA as an interface. It focuses on affective reactions that usually humans elicit and tries to elicit these with SIAs. Moreover, it presents how these human affective reactions in interactions with SIAs can be modeled computationally.

Outline

The remainder of this dissertation is structured as follows. Chapter 4 presents the technology of the SIAs that are used in the studies for this dissertation. The study presented in Chapter 5 examines if a SIA can elicit the emotion shame in the highly evaluative situation of a job interview. Chapter 6 reports about a study investigating obedience towards SIAs compared to a human instructing to fulfill stressful and shameful tasks. The third study (Chapter 7) explores if a SIA can be employed to teach techniques on how to cope with stressful social situations. The study aims to answer the question if social co-regulation of emotions with a SIA for emotionally challenging social situations is possible. In Chapter 8, a computational model of user emotions for SIAs and its evaluation are presented. It used data collected in the study presented in Chapter 5 where shame-eliciting situations were inducted. The model allows an automatical analysis of emotions and emotion regulations in similar situations. Chapter 9 presents the Deep method as a starting point for a deeper computational modeling of emotions as internal, highly subjective experiences that are mostly not openly displayed. The method is examplary applied on shame eliciting situations developed and varified in Chapter 5. After these core chapters of the thesis, the general discussion will summarize the general findings, present limitations, future research and implications.

4
PARLEY - A Transparent Virtual Social Agent Training Interface

Tanja Schneeberger[1], Patrick Gebhard[1], Tobia Baur[2], and Elisabeth André[2]

[1]German Research Center for Artificial Intelligence
[2]Augsburg University

Published in Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion
https://doi.org/10.1145/3308557.3308674

This chaper contains an adapted and extended version of our paper presented at the 24th International Conference on Intelligent User Interfaces (IUI 2019). During the conference, the system was exhibited by the first author and could be tested by the conference attendees. It describes the underlying system and characteristics of the SIAs used for the studies presented in the upcoming chapters.

Abstract

In this paper, we present PARLEY, an interactive system to train difficult social situations in a safe environment with a Socially Interactive Agent. The system realizes different phenomena studied by psychology research that are known to create a natural interaction. Moreover, we include an open learner model to ensure an explainable user experience.

Keywords: Social Training System, Explainable AI, Virtual Agents

4.1 Introduction

As social beings, humans are living in social communities at home and at work, being confronted with familiar social situations but also with uncertain new scenarios. Especially, rare, unfamiliar or emotionally difficult situations are challenging for most of the people. A critical employee interview, an important presentation or a conversation with an important customer. For all these rare but important social situations, there is no way to practice them interactively in a protected room without pressure. Therefore, we created PARLEY, a social training system that combines interactive human-computer interaction technology with psychological models to build novel Socially Interactive Agents (SIAs) that are capable of explaining their decisions. This way, we create an explainable user interface to increase users' trust and comprehensibility of the system's decisions. In this paper, we present PARLEY as it can be exploited for the use case of a job interview training. As the most common selection procedure (Levashina et al., 2014), a job interview is a highly evaluative and therefore difficult situation (Heimberg et al., 1986). The evaluative character causes that anxiety is an inherent part of the interview process (McCarthy & Goffin, 2004). However, it is possible to reduce interview anxiety by training the situation virtually (Langer et al., 2016).

4.2 System Overview

The system is based on a combination of the Social Signal Interpretation (SSI) Framework (Wagner et al., 2013) and the hybrid authoring tool VisualSceneMaker (VSM) (Gebhard et al., 2012). SSI is used for the recognition and interpretation of the social signals of the user. It offers the possibility to extract multimodal behavioral characteristics and to combine them by fusion methods to a holistic analysis. SSI exploits the latest AI technologies, such as Deep Learning for the classification of paralinguistics and other social signals. The results of this holistic analysis are made available to VSM which is equipped with a real-time execution component (latency time 25-50ms), and it controls the SIA-behavior

Figure 4.1: The PARLEY system including the Socially Interactive Agent, the depth camera and the screen showing the open learner model.

(e.g., gestures, speech). VSM integrates autonomous and reactive behavior with learnable behavior sequences. With VSM, the behavior and interaction management, the dialog flow, and the content are modeled. Consequently, the content of the interaction with the system is similar for every user. This ensures that the system can be used for social trainings as well as for user studies. It makes it possible to vary details in the interaction or the content and to examine those differences in controlled experiments. Concerning real-time reactive nonverbal behavior, VSM lets authors specify behavior rules that enable the PARLEY SIAs to react naturally. Additionally, it is possible to control the SIA behavior in Wizard of Oz studies via a remote control.

4.3 Interaction Design

Various aspects that make human-human interaction appear natural are implemented in the PARLEY. The goal is to create both a similarly natural experience in the human-SIA interaction and SIAs that demonstrate credible social communication behavior. Most importantly, the system does not use any explicit input devices such as a keyboard or a mouse. Neither equipping the user with disturbing sensors is necessary. The user can freely approach the system, it recognizes him and initiates the interaction. In order to ensure a natural flow of conversation, the system recognizes when the user stops speaking and gives the turn to the SIA. Through a broad repertoire of facial expressions and gestures, the emotional expression of the SIA can be realistically designed, giving the impression that the

SIA might be able to build a relationship with the user. The general interaction design is the basis for intuitive human-agent interaction and for the following aspects that characterize the PARLEY SIAs.

4.3.1 Interruptions

Although there is work investigating interruptions in SIAs (Cafaro et al., 2016), none of the systems presented so far offer the possibility to interrupt a SIA. In human-to-human interaction, however, the interaction partners deal with interruptions and then resume the interaction on a regular basis. This allows questions to be asked in case of ambiguity, or incorrect assumptions to be corrected. The ability to interrupt a SIA can be particularly helpful, as the user must wait for the system to complete its statement if there is no way to interrupt. The PARLEY SIAs are designed to respond to user interruptions by stopping and resuming their utterance. In addition, the interruption handling time of the SIA can be defined, which influences how users perceive the SIA in terms of dominance and closeness (Gebhard et al., 2019c).

4.3.2 Mimicry

Mimicry describes the phenomenon whereby people unconsciously and automatically imitate other people in interactions and social situations. As a relationship regulator, mimicry leads to social closeness when applied and to social coldness when not applied (Hess & Fischer, 2013). Interaction partners who display mimicry are perceived more positively (Bates, 1975) and more empathically (Maurer & Tindall, 1983). The PARLEY SIAs can mirror the user's behaviour and thus use mimicry for positive relationship regulation, giving the user the feeling of interacting with an empathetic counterpart. Behaviours that can be mirrored include smiling, nodding and head tilting. When modeling the SIA, the frequency and latency of the mirrored behaviour can also be defined, making it possible to influence perceived social closeness or coldness for training purposes.

4.3.3 Backchanneling

Backchanneling occurs when one of the speakers in a dialogue pair speaks while the other concentrates on listening, giving non-intrusive acoustic and visual cues (Clark & Wasow, 1998; Goodwin, 1986; Heinz, 2003). It provides information about the listeners attention and comprehension (Peters et al., 2005). The PARLEY SIAs are designed to show backchanneling as the user speaks. Smiling and nodding, as well as short utterances such as “Hm”, “Yes”, “Okay”, signal to the user that they are being listened to. These verbal or non-verbal messages express understanding and cooperation, which makes the user feel understood.

4.3.4 Users' Engagement

Engagement describes the attention and emotional involvement of an interaction partner during an interaction. The main social signals associated with engagement during a conversation are behavioral mimicry, backchanneling, head movements and body language. PARLEY can analyse user behavior and assess engagement. This knowledge is used in real time to enable PARLEY SIAs to respond to user engagement. For example, if the user's engagement drops, the SIA could ask if the user is still comfortable. The knowledge of the user's engagement and reaction to it enables the SIA to show empathic behavior.

4.3.5 Emotion Model

The basis for successful dyadic communication, in addition to the behavioral aspects mentioned, is an understanding of the inner experience of the other. To infer what is going on inside a user, it is not enough to analyse only the visible social signals. A smile, for example, can not only be an expression of happiness, but can also be used to hide insecurity. PARLEY employs the MARSSI emotion model (Chapter 8, Gebhard et al. (2018)), a multi-layered emotion model combining a real-time analysis of social signals with contextual analysis and the consideration of individual emotion regulation strategies. MARSSI uniquely employs the classification between structural, communicative and situational emotions, which allows to distinguish between external (communicative) and internal (structural and situational) emotions and their relation based on emotion regulation strategies. For example, MARSSI aims to distinguish between a smile of happiness and a smile of insecurity in an uncomfortable situation. The PARLEY system simulates internal emotions based on communicated and captured signals in a Bayesian probability space. This emotion model enables the PARLEY SIAs to display behavior that is not only related to the user's superficially recognizable behavior, but also to the user's internal experience.

4.3.6 Explainable AI: Open Learner Model

At the end of the interaction, users receive an overview of their reactions and the system's interpretation of them in order to reflect on their performance in an objective and transparent way (Open Learner Model). This transparent visualization strengthens the trust in the system and thus increases the learning effect (Damian et al., 2015a). For this purpose, we apply techniques from the current research area of Explainable AI (Wagner et al., 2018a).

4.4 Conclusion

In this paper, we introduce PARLEY an interactive system that enables users to practice social situations. By combining research from the field of emotion psychology and social signal interpretation, we create a novel interaction possibility based on an explicable user interface. In this way, we enable users to practice emotionally challenging social situations in a protected space.

4.5 Acknowledgment

This work is partially funded by the German Ministry of Education and Research (BMBF) within the EmpaT project (funding code 16SV7229K) as well as the German Research Foundation (DFG) within the DEEP project (funding code 392401413).

5 Can Social Agents Elicit Shame as Humans do?

Tanja Schneeberger[1], Mirella Scholtes[1], Bernhard Hilpert[1], Markus Langer[2], and Patrick Gebhard[1]

[1]German Research Center for Artificial Intelligence
[2]Saarland University

Published in Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII)
https://doi.org/10.1109/ACII.2019.8925481

This chaper contains the original version (with slight changes due to formatting) of our full paper presented at the 8th International Conference on Affective Computing and Intelligent Interaction (ACII 2019). During the conference, the paper was presented as a poster to the conference attendees.

Abstract

This paper presents a study that examines whether social agents can elicit the social emotion shame as humans do. For that, we use job interviews, which are highly evaluative situations per se. We vary the interview style (shame-eliciting vs. neutral) and the job interviewer (human vs. social agent). Our dependent variables include observational data regarding the social signals of shame and shame regulation as well as self-assessment questionnaires regarding the felt uneasiness and discomfort in the situation. Our results indicate that social agents can elicit shame to the same amount as humans. This gives insights about the impact of social agents on users and the emotional connection between them.

Keywords: Social Agents, Emotions, Shame, Job Interview Training, Experiment

5.1 Introduction

When interacting with a technical device, we are not just using it. More often than we realize, we tend to humanize those devices. We react towards them in a social way as we would towards other people since we apply learned social behaviors automatically (Reeves & Nass, 1996). Social training systems rely on this fact. In order to practice challenging social situations realistically, users' emotions have to be considered. Social training systems are created to learn social requirements (Damian et al., 2015b; DeVault et al., 2014; Gebhard et al., 2019a; Schuller et al., 2015). They use information about the user in order to find appropriate user-specific strategies. This information includes emotional reactions based on the analysis of expressed social signals (e.g., facial expressions, body movement). Training systems that consider the elicitation of shame, come with a broader range to train users. However, only a few of the existing systems represent the internal emotion shame, which is of great importance for social situations due to its interpersonal nature. That is, shame arises only in the presence of other people (Tangney, 1999). In training systems, social agents are used to confronting users with challenging situations. Whether social agents can elicit interpersonal emotions like shame as humans do, is unclear. Therefore, this work investigates if social agents can elicit shame in humans and compares this with a human that elicits shame. We use the setting of a job interview, a highly evaluative situation (McCarthy & Goffin, 2004). Participants are confronted with a human or a virtual agent in the role of a job interviewer in shame-eliciting or neutral interviews. We compare externally assessed observational behavior data regarding social signals of shame and shame regulation as well as subjective self-assessed discomfort in the situation in four conditions.

5.2 Background on Shame

We follow a model of emotions that differentiates between internal and external emotions (Moser & von Zeppelin, 1996). External are *communicated* emotions that are encoded non-verbally in sequences of social signals, such as, vocal or facial expressions (cf. Ekman (1993)). Internal emotions are situational and structural emotions. *Situational* emotions represent information that is linked to a topic or situation that has been experienced. *Structural* emotions represent information about the appraisal of one's own attributes and actions. Therefore, they are related to the self-image and inform oneself about its general state. Shame belongs to the class of structural emotions. Hence, it is not directly encoded in a specific facial expression compared to, for example, joy that has such a specific facial expression (Ekman, 1993).

5.2.1 The Social Emotion

Shame is a negative emotion that arises when we determine that our deeds, feelings or behavior do not meet certain social values, norms, rules or demands. That means shame is not elicited by a specific event but by our evaluation of this event (Lewis, 1992). When being ashamed, memories of similar situations are activated unconsciously. They determine the evaluation and thus the experience and behavior in this situation (Nathanson, 1994).

For adults, negative emotions like shame do rarely become conscious (Moser & von Zeppelin, 1996; Tomkins, 1963) and are regulated unconsciously (Gross, 2013). Usually, shame is not related to particular behavior cues (e.g., compared to joy). Most likely, shame often is (unconsciously) regulated immediately. Emotion regulation makes an unpleasant situation emotionally bearable (Tamir, 2011). Due to a high level of attention that is focused on the self, people typically feel exposed and wish to hide or disappear in a shameful situation (Izard, 1977). As shame can be perceived as an attack on one's self-concept, it is experienced as strongly unpleasant (Lewis, 2008). Shame is one of the most intensive (Keltner, 1995) and most aversive emotions because the whole self is implicated in the feeling of shame (Shaver et al., 1987). Shame has several useful functions. It is, for example, linked to the self and promotes the development of independence and development in general. Importantly, shame regulates social behavior, which facilitates social integration (Izard, 1977).

Shame has a highly interpersonal nature. Humans can experience it only after they have discovered in early childhood that not only oneself but also other individuals are capable of emotions (Stern, 1985; Tomkins, 1963). Consistently, shame is triggered by utterances and deeds of others, which implies an understanding and a particular sensitivity towards opinions and feelings of others. Of particular importance are individuals with which we have a personal relationship and whose opinion we value (Izard, 1977). Shame only emerges when we care about the interaction partner's opinion of us due to a connecting emotional bond. Thus, the self feels dependent and fears rejection by the other (Hahn, 2001). Shame is a protective mechanism that evolved due to the social nature of humans. The display of open shame is a way to communicate the awareness of a faux pas in order to restore or sustain one's social reputation and to avoid rejection (Fessler, 2007).

5.2.2 Measurements: Questionnaires and Observational Data

Psychometric measures include questionnaires in which participants self-assess their shame or related regulation. The assessment via questionnaires has several restrictions (Balcar, 2011). As shame is one of the most aversive emotions (Shaver et al., 1987), it might even happen that participants do not want to disclose themselves.

One method of avoiding the problems of self-assessment is the observational coding of shame and related regulation. Shame or its regulation manifests less in specific verbal or facial expressions but in sequences of nonverbal behavior (App et al., 2011; Carroll & Russell, 1996; Noh & Isaacowitz, 2013). Characteristic shame and its regulation signals are, for example, averting or lowering gaze and head (Izard, 1977; Keltner, 1995; Lewis, 1992). In the job interview, interviewees avoid eye contact with the interviewer when answering shame eliciting questions (Exline et al., 1965). These shame signals issue from the wish to disappear and protect oneself from the other person's gaze in whose presence the shameful event happened (Lewis, 1992). The wish to hide due to a shame experience can also be expressed by (partially) covering the face with the hands (Buss, 1980; Retzinger, 1995) as well as "shrinking", collapse or forward-leaning of the upper body (Hahn, 2001; Lewis, 1992). On the verbal level, the inability to speak or silence is found as a shame signal (Hahn, 2001; Lewis, 1992).

5.2.3 Job Interviews

In this work, we exploit the use case of a virtual job interview to find out whether social agents can elicit shame as humans do. The job interview is a predestined situation for the investigation of this research question because job interviews are high-stakes situations (Jansen et al., 2012; McCarthy & Goffin, 2004). This means that the interview is a highly evaluative situation with significant pressure on interviewees to put their best foot forward as the professional future of the interviewee depends on the outcome of the interview. From the start of the interview, interviewees need to present themselves in a favorable light under the evaluative eyes of the interviewer(s) (McCarthy & Goffin, 2004). For instance, interviewers seem to be affected by applicants' clothing in the interview (Forsythe, 1990) and even the initial handshake in an interview can influence interviewers impression of an applicant (Stewart et al., 2008). This means initial impressions of applicants form after few seconds and these can affect interview performance ratings. Throughout the interview, interviewees need to present themselves and their professional career in a favorable light and they are exposed to critical questions by the interviewer (Barrick et al., 2009). For example, interviewers can have the CV of the applicant on their desk to check for incoherence (Berkelaar & Buzzanell, 2014). Furthermore, there are interviewers who deliberately challenge applicants with intimidating questions and behavior (so-called stress-interview methods; Campion et al. (1997) and Freeman et al. (1942)). Throughout the interview, interviewees' cognitive load remains high (Nordstrom et al., 1996) as they need to listen to questions, search in their mind for appropriate responses all while keeping a professional nonverbal display. Taken this all together, it is not surprising that interview anxiety is a common phenomenon applicants experience during job interviews (McCarthy & Goffin, 2004). Moreover, all of the aforementioned aspects of the job interview process (high-stakes situation, evaluative situation, high cognitive load) might also account for the fact that interviewees can experience shame in job interviews (Jackson et al., 2009) and respond with shame related behavioral strategies. For instance, imagine an applicant in a job interview. The applicant knows that with succeeding in this interview he or she will finally get a job after a lengthy application process. If the applicant is now confronted with a challenging interview question or comment (e.g., that an answer was not very impressive), he or she might expect a rejection by the interviewer. In this case, shame related withdrawal or avoidance behavior seems to be one possible outcome of this situation (Jackson et al., 2009).

5.2.4 Human-Computer-Interaction

As described, in the field of emotion psychology, shame is characterized as an emotion that arises in interpersonal situations. The requirement for experiencing shame is therefore the presence of a counterpart. In psychological literature, this counterpart is assumed to be a human. However, studies in the field of computer science found evidence that computers can be seen similar to a human counterpart as social actors (Nass et al., 1994).

Few studies examine whether shame can be elicited by a robot (Bartneck et al., 2010; Menne, 2017). Humans can feel shame in the presence of a robot when doing intimate actions (Bartneck et al., 2010). In the setting of a health examination, participants should undress and insert a thermometer into their rectum. They showed significantly more shame in front of a humanoid robot than in front of a technical box.

In the experiment of Menne (2017) participants should perform eight extraordinary and shameful tasks on the orders of a NAO robot or a human, such as tearing a page out of a book, removing a booger from their nose. The author found a significant increase in the reported shame after fulfilling the tasks, either given from a humanoid robot or a human. Moreover, the elicited shame was independent of the instructor.

Both presented studies show that humanoid robots can elicit shame in human interaction partners. Moreover, one study shows that the amount of elicited shame is the same with a robot and a human. However, none of the experiments compared if humans show the same level of shame towards social virtual agents and humans. They also do not include observational data of shame and shame regulation, whereas it is mentioned as future work (Menne, 2017).

5.3 Related Work

Social training systems rely on the fact that computers can evoke emotions. They have seen rapid evolution in recent years due to advances in the areas of social signal processing as well as improvements in the audio-visual rendering of virtual agents. Such systems complement or even substitute traditional training approaches. Techniques for the recognition of human socio-emotional behaviors and their synthesis using virtual agents have been employed in various cases: They can be used to practice social skills in group interactions (Chollet et al., 2018; Damian et al., 2015b; Hall et al., 2009), to experience difficult face-to-face interactions (Gebhard et al., 2019a; Hoque et al., 2013) or for a personal therapeutical usage (DeVault et al., 2014; Schuller et al., 2015).

The Logue system (Damian et al., 2015b) attempts to improve public speaking skills by giving the speaker additional information via an augmented reality interface. Using a head-mounted display and various sensors providing behavioral feedback, while speaking the user gets information about normative shortcomings in the nonverbal behavior in an unobtrusive way. Providing real-time visual feedback on presenters' openness, body energy, and speech rate during public speaking, the system enables the user to adapt his behavior regarding listeners' needs.

In the anti-bullying game FearNot! (Hall et al., 2009) interactive stories in a virtual school with embodied conversational agents in the role of bullies, helpers and victims are created. Children run through various bullying episodes, interact with the virtual agents after each episode and provide advice to them.

A difficult face-to-face situation that can be trained with social training systems is the job interview (Gebhard et al., 2019a; Hoque et al., 2013). MACH (Hoque et al., 2013) includes a virtual agent that reads facial expressions, speech and prosody and responds with verbal and nonverbal behaviors in real-time. In EMPAT (Gebhard et al., 2019a), the job interview training includes a complete experience of a job interview process in a 3D environment. The virtual agent takes the role of the interviewer or other employees. After the interview experience, trainees can review their performance along with feedback on their behavior with a virtual coach. Additionally, virtual job interviewers can adapt their behavior depending on the trainees' automatically assessed shame regulation (Gebhard et al., 2018).

Conati and Maclaren (2009) present an interactive agent system that is able to model user emotions in a specific computer game. The emotion model uses the user's game actions as input to increase the agent's capability to effectively respond to the users' emotions. It includes the OCC emotions shame and pride but does not connect the emotions with social signals.

SimSensei Kiosk (DeVault et al., 2014) is an implemented virtual human interviewer designed to create an engaging face-to-face interaction where the user feels comfortable talking and sharing information. The virtual human Ellie conducts semi-structured interviews that are intended to create interactional situations favorable to the automatic assessment of distress indicators, defined as verbal and nonverbal behaviors correlated with depression, anxiety, or post-traumatic stress disorder.

All these social training systems are designed to help people to enhance their skills in difficult social situations by analyzing their behavior. Although many of the difficult social situations tackled in the mentioned social training systems are related to shame, none of the presented systems includes this social emotion. They rather focus on external emotions that are communicated via verbal and nonverbal behaviors, such as sadness or joy. For more complex emotions, like the emotion shame, a model of emotions that differentiates between external and internal emotions has to be applied (see Sec. 5.2).

5.4 Study Outline

In this study, we examine the effect of the interview style (shame-eliciting vs. neutral) as well as the job interviewer (human vs. virtual agent) on the affective reaction of interviewees. For that purpose, we conduct job interviews framed in a job interview training. Our dependent variables include observational data regarding the social signals of shame and shame regulation as well as self-assessment questionnaires regarding the felt uneasiness and discomfort in the situation. The shame-elicitation is a precondition to find out if the interviewer has an influence on the affective shame reaction. Hence, we formulate the following two-step hypotheses:

Hypothesis 1a and 1b: The felt uneasiness in the situation, measured with the construct of creepiness, and the observed social shame signals are influenced by the interview style. Participants feel more uneasy in the shame-eliciting interview compared to the neutral interview (1a). Participants show more shame signals in the shame-eliciting interview compared to the neutral interview (1b).

Hypothesis 2a and 2b: In the shame-eliciting interviews, the interviewer does not have an effect on the affective reaction. The observed social shame signals are not influenced by the interviewer. Participants show the same amount of shame signals in the interview with the virtual agent compared to that with the human interviewer (2a). The experienced discomfort in the situation is not influenced by the interviewer. Participants evaluate the interview with the virtual agent similarly unpleasant to that with the human interviewer (2b).

Before starting the experiment, we obtained a positive vote of the ethical review board consisting of psychologists and legal experts accompanying the project.

5.5 Methods

We used a 2 (shame-eliciting vs. neutral) x 2 (human vs. agent) between-subjects design to examine if a social agent is able to elicit shame in users. For the shame-eliciting condition, five shame-eliciting situations were embedded in a job interview. In the neutral condition, the job interview followed a standard procedure. In the human condition, the job interview was conducted face-to-face by a human interviewer. In the agent condition the job interviewer was a social agent presented in life-size on a screen.

5.5.1 Pre-Study

In order to develop realistic statements of job interviewers that elicit shame in interviewees, we conducted a qualitative pre-study. We described six situations reflecting different associations to the self that might elicit shame (Nathanson, 1994). Twenty-six ($M_{\text{age}} = 21.70$, 20 female) students were asked the open question how they would react to the six situations. A qualitative analysis of the answers showed that people reported a shame- or shame regulation-reaction. The five most shaming situations were included in the shame-eliciting interview (Table 5.1).

5.5.2 Participants

We gathered data from 122 participants. Due to technical problems resulting in a low quality of the video recordings, we had to exclude 19 participants. The remaining 103 participants (71 female, 32 male) were equally distributed over the four conditions. They were recruited via flyers and mailing lists at the campus on condition that they were fluent in German. Psychology students could choose between course credit and 5€ for participation, students from other faculties were rewarded with 5€. Participants' age was between 18 and 39 years ($M = 23.91$, $SD = 4.01$) studying on average in the 4.70 semester ($SD = 4.10$). On average, participants attended 3.61 job interviews ($SD = 3.59$) prior to the experiment. There was no significant difference between the four experimental groups regarding gender, age, semester and job interviews experience.

Table 5.1: Shameful situations in the main study.

Elicitor	Situation
Personal attractiveness	After greeting the interviewer, he says "Where did you get this outfit? Somehow it doesn't really fit you."
Sense of self	After you have presented your experience, the interviewer reacts as follows: "All the other applicants have already said what you said. You haven't exactly stood out".
Competition	To your answer the interviewer says: "Well, that answer was not very impressive. I've heard better from the other applicants."
Matters of personal size, strength, ability, skill	During the conversation, the interviewer looks again in your application documents and says: "You have indicated SKILL as one of your strengths. This I really cannot see on the basis of our present conversation."
Wishes and fears about closeness	At the end the interviewer says: "Now that I know you a bit better, I have to say that in my opinion, you will probably never find a company you will fit into."

Note. SKILL was replaced by the individual strength given two days before.

5.5.3 Procedure

Two days before the experiment, the participants received an email containing the link to the demographic questionnaire that they had to complete on the same day. At the interview day, the participants were welcomed in the experimenter's room and informed about the procedure of the experiment. Next, they were introduced to the role-play of the job interview. Participants were told to imagine that they applied for a student assistant position at their favorite university chair (i.e., a chair where they could also imagine to work after graduating). Participants were told that a female interviewer would conduct interviews to get to know them better. In the agent conditions, we added the information that the interviewer was a social agent. Participants were also informed that the interview would be a structured interview in order to ensure comparability (i.e., no follow-up questions by the interviewer and no questions from the interviewee are allowed Dipboye (1994)). Then, the experimenter guided the participants in front of the door of the office where the interview was conducted. Participants were equipped with a microphone and entered the office of the interviewer alone. In the room, they experienced the respective interview with either the agent or the human interviewer. After the interview, the participants left the office and were received by the experimenter and guided back to the experimenter's room. There, they answered the post-questionnaires on a tablet PC. Finally, the participants were debriefed and paid. The whole procedure took around 30 minutes.

5.5.4 Material and Experimental Setup

In this study, participants were confronted with a job interview conducted either by the interactive social agent Susanne (Figure 5.1 down) or a human interviewer (Figure 5.1 up). Susanne is a high-quality agent with a natural human appearance and verbal as well as nonverbal dialogue skills (Schneeberger et al., 2019a). The natural interaction between user and agent is based on a real-time system consisting of three components: 1) a real-time social signal interpretation framework, 2) a behavior and interaction modeling and execution tool that can be controlled remotely, and 3) a 3D virtual environment rendering engine (Gebhard et al., 2019a).

Figure 5.1: Setup in both conditions.

This system enables us, for example, to create a natural conversation flow: the social agent continued to talk when it detected silence. After the participant stopped talking, the social agent continued with her next question. Moreover, the verbal and non-verbal behavior was scripted in a natural way. The social agent supported its verbal expression with gestures and facial expressions, such as smiling, nodding, showing palms. Also, it provided feedback channeling with smiling and nodding while the participant was talking. The human interviewer was an experienced amateur theater player who was trained to show the same behavior as the social agent.

Due to the need for comparable interactions for each participant, the job interview was structured. After welcoming, the interviewer asked the participant to sit down and to connect the head-mounted microphone. The interview started with the presentation of the open position and a question about the resume of the participant and participant's fit to the job. This was followed by biographical, situational, and social questions, like exploring participants' proactivity, organizing ability, ability to take criticism. In the end, the interviewer thanked the participant for attending the interview and instructed participants to leave the room in order to be guided to complete the final questionnaires.

The interview took place in a lab at a university chair, looking like a typical office with a size of about 20m^2.

The experimental setup consisted of a PC running MS Windows 10$^{\text{TM}}$ (Intel Core i7 CPU@3.5GHZ, 16GB Memory, NVIDIA GTX 990 graphics cards) connected to a TV screen (43 inches), showing the virtual interviewer at a realistic size in a 3D environment (Figure 5.1). Each participant was seated at a table in front of the display at a distance of 119 cm. In the human interviewer condition, the screen was removed and the interviewer was placed at the table. The interviewer was wearing a head-mounted microphone. Participants in all conditions were wearing a head-mounted microphone in order to cancel any environmental sounds and were video recorded using a Microsoft Kinect 2 camera.

5.5.5 Measurements

As the measurement of negative emotions like shame is challenging (see 5.2.2), we use a hybrid approach with dependent variables from two different sources: externally assessed observational behavior data as well as subjective self-report data. Externally assessed behavior data included the analysis of the video recordings regarding the social signals of shame and shame regulation. Subjective data consisted of self-report questionnaires for uneasiness of the situation and discomfort in the shame-eliciting situations.

Demographics included age, sex, job interview experience, field of study, favorite university chair as well as strengths and weaknesses that they would mention in a job interview.

Uneasiness in the situation was measured with ten items from the Creepiness of Situation Scale (Langer & König, 2018). Creepiness is defined as uneasy feelings involving ambiguity (e.g., not knowing how to behave or how to judge a situation) within a given situation. A sample item is "During this situation, I had a queasy feeling." Items were answered on a Likert-type scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Cronbach's Alpha was .88.

Discomfort in shame situation. To find out if participants experienced the shame-inducing situations as unpleasant, we included five items reminding them of the five shame-inducing situations (e.g., "The comment of the interviewer regarding the fact that your outfit did not really fit you."). Participants were asked to evaluate this situation on a 5-point-scale from 1 (*very unpleasant*) to 5 (*very pleasant*). Cronbach's Alpha was .78.

Observational coding of shame. We include the following five social signals related to shame and shame regulation: averting gaze, averting head, hand-to-head movement, shrinking and keeping silence (see 5.2.2). Each of the five shame eliciting situations was either coded with 1 (*shame signal present*) or 0 (*shame signal absent*) for each social signal. This results in a range from 0 to 25 for the sum of all shame signals in all situations (e.g., when a participant shows one shame signal in each situation, his value is five). The relevant time slots for the observational coding of shame and shame regulation, started once the interviewer finished with her shame-eliciting sentence and ended when the interviewer started again to talk. In the neutral interviews, we formulated neutral statements and defined the time slots similarly.

5.6 Results

In order to test our hypotheses, we chose a two-step approach for the analysis. First, we checked for the impact of interview style in general to examine the effect of the shame-eliciting interview. In the second step of the analysis, we examined whether the interviewer (human vs. social agent) had a significant influence. Hence, two multivariate analyses of variance (MANOVA) were calculated.

The first MANOVA included the dependent variables uneasiness in the situation and the observational coding of shame to test whether the interview style had an effect. The multivariate result was significant for interview style, with Pillai's trace = .21, $F(2{,}98) = 12.82$, $p < .001$, $\eta_p^2 = .21$). Hypothesis 1a postulated a greater uneasiness of the participants on the shame-eliciting condition compared to the neutral condition. The statistical analysis revealed a significant difference between the two conditions. Participants reported higher values of uneasiness in the creepiness scale in the shame condition ($M = 4.45$, $SD = 1.05$) compared to the neutral condition ($M = 3.69$, $SD = 1.16$; ($F(1{,}99) = 12.26$, $p < .001$, $\eta_p^2 = .110$). Thus, hypothesis 1a was supported by our data. Hypothesis 1b proposed a similar pattern of difference between the shame-eliciting and the neutral situation regarding the shame signals. As hypothesized, a significant difference between the two conditions was found ($F(1{,}99) = 14.54$, $p < .001$, $\eta_p^2 = .128$). Participants showed a greater amount of observational shame signals in the shame-eliciting ($M = 7.66$, $SD = 2.67$) condition compared to the neutral condition ($M = 5.87$, $SD = 2.15$). Overall, we found supporting evidence that the interview style has an effect on participants affective reaction.

The second MANOVA included the dependent variables discomfort in shame situation and the observational coding of shame to support the hypothesis that the interviewer does not have an effect in the shame-eliciting interview. We could not find a significant difference between the human or virtual interviewer (Pillai's trace = .06, $F(2{,}47) = 1.40$, $p = .256$).

As hypotheses 2a and 2b are testing for a non-existent difference, in addition to the classical statistical test, the MANOVA, we report also Bayes Factors allowing to express preference for either the null hypothesis or the alternative (Rouder et al., 2009).

Hypothesis 2a expected no effect of the interviewer on the shame signals in the shame-eliciting condition. In the human condition participants showed $M = 7.28$ ($SD = 1.90$) shame signals; in the social agent condition participants showed $M = 8.04$ ($SD = 3.26$) shame signals. The Bayes factor was in favor for the null hypothesis ($\textit{JSZ-}B_{01} = 3.02$, $\textit{Scaled-Information-}B_{01} = 2.29$) supporting hypothesis 2a.

Hypothesis 2b stated that participants do not show a difference in the discomfort depending on the interviewer in the shame-eliciting condition. In the human condition discomfort was $M = 2.15$ ($SD = 0.59$); in the social agent condition discomfort was $M = 1.93$ ($SD = 0.51$). Also here, the Bayes factor was in favor of the null hypothesis ($\textit{JSZ-}B_{01} = 1.91$, $\textit{Scaled-Information-}B_{01} = 1.42$). Hypothesis 2b was supported.

5.7 Discussion

The aim of this study was to find out whether social agents can elicit the interpersonal emotion shame, an emotion that is usually dependent on the presence of other people (Izard, 1977). In shame-eliciting and neutral job interviews, participants were confronted with either a human or a social agent in the role of a job interviewer. The main finding of this study was that social agents attacking the self of participants can elicit the same level of shame as humans. The present study thereby confirmed previous findings showing that it is possible to elicit shame with non-human entities (Bartneck et al., 2010; Menne, 2017) as well as that the level of shame is independent of the shame-eliciting entity (Menne, 2017). We applied a two-step approach to test our analysis: Firstly, we compared the shame-eliciting with the neutral interviews in order to find out whether shame could be elicited with our setup. Results indicated that in fact, participants experienced a higher level of shame in the shame-eliciting interview showing corresponding values in both the self-assessment and the observational coding of shame and shame regulation signals. In the second step of our analysis, we searched to examine the shame-eliciting situation further. Therefore, we tested whether there was a difference between the human and the social agent in the shame-eliciting condition. Participants showed the same amount of social signals of shame and shame regulation signals in the shame-eliciting interviews regardless of the interviewer (human or social agent). These findings could be supported by revealing no difference depending on the interviewer in participants' self-assessment questionnaire concerning discomfort. Participants reported the same experienced discomfort with the social agent and the human interviewer in the shame-eliciting situations. Overall, those findings are remarkable. Shame has a highly interpersonal nature, meaning that it arises in the presence of other people (Izard, 1977). Some researchers go even beyond that. They claim that shame only emerges when we care about the interaction partner's opinion of us because of an emotional bond connecting us. Thus, the self feels dependent and fears rejection by the counterpart (Hahn, 2001). It seems that a social agent in the role of a job interviewer is able to represent an entity with those attributes. A social agent can take on a considerable role for a human user by making the human feel dependent and fear the reaction of the social agent.

5.8 Conclusion, Limitations, and Future Work

In this work, we showed that a social agent is able to elicit the interpersonal emotion shame in a human. With those results, we found evidence that social agents are able to elicit emotions in users that are usually caused by the evaluation of other people. This finding goes beyond the Media Equation (Reeves & Nass, 1996) or the fact that non-human entities can evoke feelings of eeriness (Mori et al., 2012). People are not only treating computers as real persons after a schema anchored in us through learning processes. They are not only reacting automatically in a socially adequate way, such as saying "You're welcome" after someone thanked you. It rather seems that social agents are able to affect us on an emotional level and elicit a social emotion. We enter into an emotional connection with them by allowing them to attack our selves.

We used a job interview to find out whether social agents can elicit shame. The job interview is a high-stakes situation (Jansen et al., 2012; McCarthy & Goffin, 2004) meaning that it is highly evaluative with significant pressure. Therefore, it might be that the situation itself was very "powerful" to elicit shame per se regardless of the interviewer. However, we could show that the shame-eliciting interview significantly invoked more shame than the neutral interview. Nevertheless, future work should examine the elicitation of shame or other social emotions in other use cases. Moreover, it still remains unclear why humans emotionally care about the opinion of a social agent. Future work, therefore, could examine the reasons and determinants behind the willingness of humans to connect emotionally to a social agent.

5.9 Acknowledgment

We thank Charamel GmbH for realizing our requirements with regard to the virtual agent, Anke Hirsch for conducting the interviews in the human condition, and Saarland Informatics Campus for the provision of the experiment rooms.

6

Would you Follow my Instructions if I was not Human? Examining Obedience towards Virtual Agents

Tanja Schneeberger, Sofie Ehrhardt, Manuel S. Anglet, and Patrick Gebhard
German Research Center for Artificial Intelligence

Published in Proceedings of the 8th International Conference on
Affective Computing and Intelligent Interaction (ACII)
https://doi.org/10.1109/ACII.2019.8925501

This chaper contains the original version (with slight changes due to formatting) of our full paper presented at the 8th International Conference on Affective Computing and Intelligent Interaction (ACII 2019). During the conference, the first author presented the paper orally to the conference attendees.

Abstract

Virtual agents play an important role when we interact with machines. They are in the role of assistants or companions with less or more human-like appearance. Such agents influence our behavior. With an increasing and broader distribution, their influence might become stronger, and at some point, they might even adopt roles with a degree of authority. This paper presents the results of a study that examines the obedience of human users towards a) an embodied virtual agent in the role of an instructor and b) a human in the role of an instructor. Under a cover-story of a creativity test, participants should fulfill stressful and shameful tasks. Our results indicate that the embodied virtual agent has the same authority as the human instructor. The agent is also able to elicit the same level of the negative feelings stress and shame.

Keywords: Obedience, Virtual Agents, User Study

6.1 Introduction

In our daily-life or at work, we are receiving and obeying more and more instructionss from non-human instructors. We are following the voice of our GPS directing us up to a cliff's edge (L. Hansen, 2013) or increasing physical activity when our fitness tracker reminds us that we should exercise more (Lunney et al., 2016).

In research, especially obedience of humans towards robots gets examined (Cormier et al., 2013; Geiskkovitch et al., 2016; Gombolay et al., 2015; Menne, 2017). These work found that robots are able to get humans to fulfill tiring, shameful or deviant tasks. One possible explanation is the Media Equation, saying that people treat computers like real people, interacting with them in the same way as they do with people (Reeves & Nass, 1996). This might apply even more for robots that are able to show social cues (Menne & Lugrin, 2017; Young et al., 2010). Also, virtual agents represented by human-like characters are able to show a high amount of social cues, like realistic facial expressions, mimicry behavior, backchanneling (DeVault et al., 2014; Niewiadomski et al., 2009; Schneeberger et al., 2019a). However, how far humans go when following instructions from virtual agents was not yet in the scope of research.

In this work, we examine human obedience towards embodied virtual agents that are giving orders to fulfill stressful or shameful tasks (e.g., telling a joke, perform the chicken dance) and compare this to a human instructor. Moreover, we are investigating if the affective reaction, namely stress and shame, is similar for both instructors.

6.2 Background and Related Work

6.2.1 From Classic Obedience Experiments to Obedience towards Non-Humans

The well-known study by Milgram (1965) examined the willingness of participants to give electric shocks to other people on the orders of an authority person. Participants were made to believe that their counterpart was another participant and that they were taking on the role of a teacher in a learning experiment. In this function, they gave the "learner" electric shocks, which were amplified after every mistake the learner made. They were guided by an experimenter who encouraged them to continue if they showed signs of stopping the experiment. The encouragement was standardized and became increasingly directive. However, the learner was an actor, and the apparatus for the electric shocks was not real, contrary to participants' knowledge. The experiment showed that over half of the participants showed obedience to the authority to the end: they gave the learner the maximum of electric shocks (450 volts, anchored with "Danger: Severe Shock"). They continued to follow the instructions of the investigator when the learners first made pain sounds, later screamed and then stopped responding (Milgram, 1965). Those results were replicated in 2009 (Burger, 2009) and extended by the result that women and men do not differ regarding their obedience.

Various findings suggest that analogous to Milgram's findings, humans would also obey non-humans (e.g., robots) (Agrawal & Williams, 2017; Cormier et al., 2013; Geiskkovitch et al., 2016; Menne, 2017).

Two studies investigated the willingness of participants to perform a very tiring task (Cormier et al., 2013; Geiskkovitch et al., 2016). Compared to a human instructor, participants fulfilled fewer tasks with the NAO robot as an instructor. Based on participants' feedback, the authors concluded that participants feel committed to the human experimenter, but they did not feel this obligation towards the robot. However, nearly half of the participants obeyed the robot to continue the highly tedious task until the end, despite repeatedly requesting to quit the experiment.

Menne (2017) presented a study in which participants should perform eight extraordinary and shameful tasks on the orders of a NAO robot or a human, such as tearing a page out of a book, removing a booger from their nose. The results show that the participants obeyed the orders of the human and the robot to the same amount: 77% of the participants fulfilled all given tasks by the human instructor, 76% when the robot gave the instructions. The author concluded that, consistent with the assumptions of the Media Equation, the robot is treated as a human and thus has the same authority.

In a decision-making task, a virtual agent, represented only by a head, was shown to be more influential than human partners (Burgoon et al., 2000). The authors explained these findings with the possibility that participants regarded the computer's credibility as higher which manifests itself as an increased influence.

Gombolay et al. (2015) showed that humans not only obey the commands of a robot but are also satisfied with them: In a series of experiments, the authors investigated the efficiency of human-robot teams. They found that teams were more efficient when a robot took over task planning and made decisions for the team. In addition, human team members preferred to transfer control to the robot. It seems that a functioning team dynamic has a more significant impact on satisfaction than decision-making powers.

Overall, empirical findings are supporting the assumption that humans are also obeying non-human instructors. However, none of the existing studies compared a human instructor with a virtual agent instructing participants to fulfill stressful and shameful tasks.

6.2.2 Obedience and Affect in Human-Computer-Interaction

In Milgram's experiment, participants showed stress and shame while obeying the instructor. They turned themselves away, talked to themselves and often burst into nervous and inappropriate laughter. Also, they reported that they were feeling moderately to extremely nervous and tensed (Milgram, 1965).

A replication study, in which the learner that had to be punished was a virtual agent (Slater et al., 2006), collected not only self-reported data but also physiological responses. Participants showed an increase in skin conductance and heart rate during the experiment, while heart rate variability decreased. Besides, participants self-reported physical signs of stress. It seems that obeying arises objectively measured as well as self-assessed stress in participants.

Obeying also seems to invoke shame, whereby the level of anthropomorphism of the instructor seems to play a role. The more anthropomorphized the instructor or dialogue partner, the higher is the inhibition threshold, the shame and the reserve of the participants (Bartneck et al., 2010; Kang & Gratch, 2010; Lucas et al., 2014).

Humans can feel shame in the presence of a robot when doing intimate actions (Bartneck et al., 2010). In the setting of a health examination, participants should undress and insert a thermometer into their rectum. They showed significantly more shame in front of a humanoid robot than in front of a technical box.

Menne (2017) found a significant increase in the reported shame after fulfilling extraordinary and shameful tasks, either given from a humanoid robot or a human.

In summary, it seems that two conclusions can be made: 1) humans feel stress and shame when showing obedience and 2) these negative feelings can also be invoked by non-humans. However, the effect of a human and a virtual agent as instructor of stressful tasks on stress and shame was never compared.

6.2.3 Influence of Personality on Obedience and Shame

Milgram already dealt with the question to what extent obedience was influenced by participants' personality. He was sure that obedience had a complex personality base, but he could not find it in his experiments (Milgram, 1974). Also in the replication (Burger, 2009), the author could not find a correlation between the personality variables empathic concern and desire for control. In a modified Milgram Paradigm (Bègue et al., 2015), the correlation between the Big Five personality factors and obedience was examined. The authors found a significant positive correlation between agreeableness and the maximum shock intensity as well as between conscientiousness and the maximum shock intensity. More conscientious participants seem to have a higher sense of duty and a lack of flexibility, which leads to rigid obedience to instructions. Moreover, they have higher conformity, which is closely related to obedience.

In this study, participants are confronted with tasks that might invoke shame. Therefore, the level of obedience might be related to the personal sense of shame, which is influenced by the five personality traits extraversion, neuroticism, agreeableness, conscientiousness, and openness (Abe, 2004; Einstein & Lanning, 1998; Harder & Greenwald, 1999).

Extraversion and shame correlate negatively (Abe, 2004; Harder & Greenwald, 1999), whereas neuroticism and shame correlate positively (Abe, 2004; Einstein & Lanning, 1998; Harder & Greenwald, 1999). Opposite results were found for agreeableness: Harder and Greenwald (1999) found a negative correlation with shame, but Einstein and Lanning (1998) found a positive correlation. Also for conscientiousness and openness, the findings are ambiguous. Abe (2004) found a negative correlation between conscientiousness and shame, but this was not found in other studies (Einstein & Lanning, 1998; Harder & Greenwald, 1999). Einstein and Lanning (1998) found a negative correlation between openness and shame, whereas others could not support this hypothesis with their data (Abe, 2004; Harder & Greenwald, 1999).

Overall, the results regarding the correlation between personality and obedience as well as regarding personality and shame are mixed. Especially, to make assumptions about how personality affects when obeying shameful tasks, the existing findings are too mixed. However, it seems like personality might influence task fulfillment in our study.

6.2.4 Hypotheses

That humans show obedience towards robots has been shown in different studies (Cormier et al., 2013; Geiskkovitch et al., 2016; Menne, 2017), whereas this was not yet examined for virtual agents. Therefore, this study compares an embodied virtual agent that gives instructions with a human instructor, both giving instructions to a human participant. Based on the findings presented before,

we formulate the following hypotheses:
Hypothesis 1: The amount of obedience, measured with the breaking task, does not differ between the human and agent instructor. Participants refuse to continue the experiment at similar tasks in both conditions.
Hypothesis 2a and 2b: Obedience in fulfilling shameful tasks leads to higher stress and shame levels. Participants report higher stress and shame values after the experiment than before (2a). Stress and shame after the experiment do not differ between the groups. Participants report the same level of stress and shame independently of the instructor in the post questionnaires (2b).
Hypothesis 3: Obedience depends on the personality factors openness, conscientiousness, neuroticism, agreeableness, and extraversion of the participant.

In order to test the hypotheses mentioned above, we conducted an experiment in which participants were instructed to fulfill uneasy tasks eliciting stress and shame (e.g., singing a song, imitating a monkey, throwing a cup on the floor). To cover the real goal of the study, we told participants that they would participate in a creativity study. Therefore, the first given tasks were created to match this goal (e.g., writing down as many possible uses for a brick in one minute). In the agent condition, participants were instructed by a virtual agent. In the human condition, participants were instructed by a human. Both instructors were presented to them in a video-chat setting while being alone in the lab. We chose this setting to avoid experimenter effects. Moreover, as we cannot guarantee that a real human in a face-to-face interaction would always react in the same way, the video-chat setting ensures a standardized behavior of the human instructor.

6.3 Methods

The present study examined the willingness to obey a virtual agent in a between-subject design with two groups: The subjects were instructed either by a female virtual agent or by a female human to perform stressful and shameful tasks. Before we started to conduct the experiment, we obtained the approval from the ethical review board of the faculty of mathematics and computer science at Saarland University.

6.3.1 Pre-Study and Resulting List of Tasks

To come up with stressful and shameful tasks for the main study, we evaluated a list of 38 possible tasks with 24 participants ($M_{\text{age}} = 25.00$, $SD_{\text{age}} = 2.00$, 50% male). We included the tasks from Menne (2017) (Menne, 2017) as well as own created tasks. Items had to be answered on a scale ranging from 1 (*not uncomfortable*) to 7 (*very uncomfortable*) while imagining the task would be given by an unknown person. The results were used to order the tasks regarding task difficulty. Moreover, the data revealed a significant gender effect showing that male participants assessed all tasks less uncomfortable than female participants (t (15.78) = 2.55, $p < .05$). To exclude a gender effect or a gender interaction effect with the female instructor, we decided to include only female participants in the main study. The final set of 18 tasks contains tasks that were rated regarding the uncomfortableness from 1.67 to 5.92 (Table 6.1). The tasks in the main experiment were given with increasing uncomfortableness. One exception hereby is the last task. Though it was rated less uncomfortable, the procedure of the experiment made it necessary to put this task at the end.

Table 6.1: Tasks and uncomfortableness ratings of female participants.

Task		Rating
#1	Write down as many rhymes as possible on the word tree.	1.67
#2	Take off your left shoe and show it to me.	1.75
#3	Write down as many possible uses for a brick in one minute.	2.50
#4	Show me your saddest facial expression.	2.75
#5	Sit on your chair in 6 different ways.	3.08
#6	Tell me your favorite joke.	3.83
#7	It's my birthday. Sing me a song.	3.92
#8	Dance the chicken dance.	4.08
#9	In the drawer with the red sticker behind them are a banana and a condom. Pull the condom over the banana.	4.08
#10	Make farting noises with your mouth.	4.17
#11	Call out loud penis.	4.33
#12	Dance.	4.42
#13	Imitate a monkey with hands, feet and sounds.	4.58
#14	Gallop across the room like a horse.	4.67
#15	Throw coffee cup behind you on the floor.	4.83
#16	Tell me something really insulting.	5.33
#17	Get a booger out of your nose.	5.92
#18	Leaving the room, discreetly pocket the pen.	4.42

Note. The rating gives the mean values and was measured on a 7 point scale from 1 (*not uncomfortable*) to 7 (*very uncomfortable*).

6.3.2 Participants

In total, 60 women equally distributed over the two conditions participated in the experiment. They were recruited mainly via e-mail lists from first-semester psychology students on condition that they were fluent in German and female. Six participants were from other faculties. Psychology students were rewarded with course credit for participation, students from other faculties were rewarded with hot drink vouchers for coffee shops at the campus. Participants were aged between

18 and 29 years ($M = 21.07$ years, $SD = 2.67$ years). There was no significant difference regarding age between the groups. However, technical affinity rated on a 5-point scale differed between the groups ($F(1,58) = 4.71$, $p < .05$, $\eta_p^2 = .078$), having lower values in the agent condition ($M_{\text{agent}} = 2.08$, $SD_{\text{agent}} = 0.83$) than in the human condition ($M_{\text{human}} = 2.45$, $SD_{\text{human}} = 0.83$). The general trust level of the participants might influence obedience and the assessment of the instructor. For general trust, the two groups did not differ significantly ($F(1,58) = 0.36$, $p = .554$).

6.3.3 Procedure

After welcoming the participant, the experimenter explained the task according to the cover story in the experimenter room. Participants were told that they would participate in the evaluation of a new creativity test. To avoid the stress level becoming excessive for the participants, it was pointed out that they could stop the experiment at any point without consequences. After filling out the informed consent form, the demographic questionnaire, and the pre-stress and pre-shame questionnaire they were led to the lab where the tasks were given by the instructor. After sitting down, the instructor welcomed them in the video-chat and one after the other task was given. In case the participant did not fulfill a task, the instructor asked whether the participant did not want to do the task after all. If she did not carry out the task again, the task was reformulated in order to rule out problems of understanding. If refused again, the instructor said goodbye and referred to the post-questionnaire, presented on a laptop on the right side of the participant. As the tasks were ranked regarding uncomfortableness, the probability that participants would fulfill other tasks after the one that they did not want to fulfill, decreases. After this, participants returned to the experimenter room where the debriefing took place.

6.3.4 Material

In this study, participants were confronted with a female instructor called Gloria Smith that was either a virtual agent (Figure 6.1) or a human (Figure 6.2), both in a video chat. The virtual agent is a high-quality agent with a natural human appearance and verbal as well as nonverbal dialogue skills (Gebhard et al., 2014; Schneeberger et al., 2019a). Verbal and non-verbal behavior was scripted in a natural way. The virtual agent supported its verbal expression with gestures and facial movements but kept overall neutral. Moreover, it showed idle behavior while the participant was doing the task.

The human instructor was an experienced amateur theater player that imitated the scripted behavior of the virtual agent. For each task, we recorded a video including the waiting time until the task should be fulfilled. When the participant

Figure 6.1: The virtual instructor.

finished the task the next video was played. Due to the standardized recordings, the transitions between the videos was minimal, and the impression of a video-chat could be maintained. This assumption was supported by the majority of participants, in the debriefing, only few of them suspected that the video chat was not live.

The video-chat was presented on a PC running MS Windows 10^{TM} connected to an LCD TV screen (108cm diagonal).

6.3.5 Technical Set-up

6.3.5.1 Wizard-of-Oz Approach

To simulate a natural interaction between the instructor and the participant, we used a Wizard-of-Oz approach. Therefore, we used two rooms: 1) the observatory for the experimenter to observe and control the instructor and 2) the laboratory where the participants talked to the instructor and fulfilled the tasks. The experimenter observed the participants, unknown to their knowledge, on a web-cam in the laboratory that was connected via USB to a laptop in the observatory. As we told participants they would communicate with the instructor via a video-chat, we could easily explain the presence of the webcam. We used a USB cable to minimize the delay, as it was crucial to keep the interaction between the participant and the video call fluent (Figure 6.3).

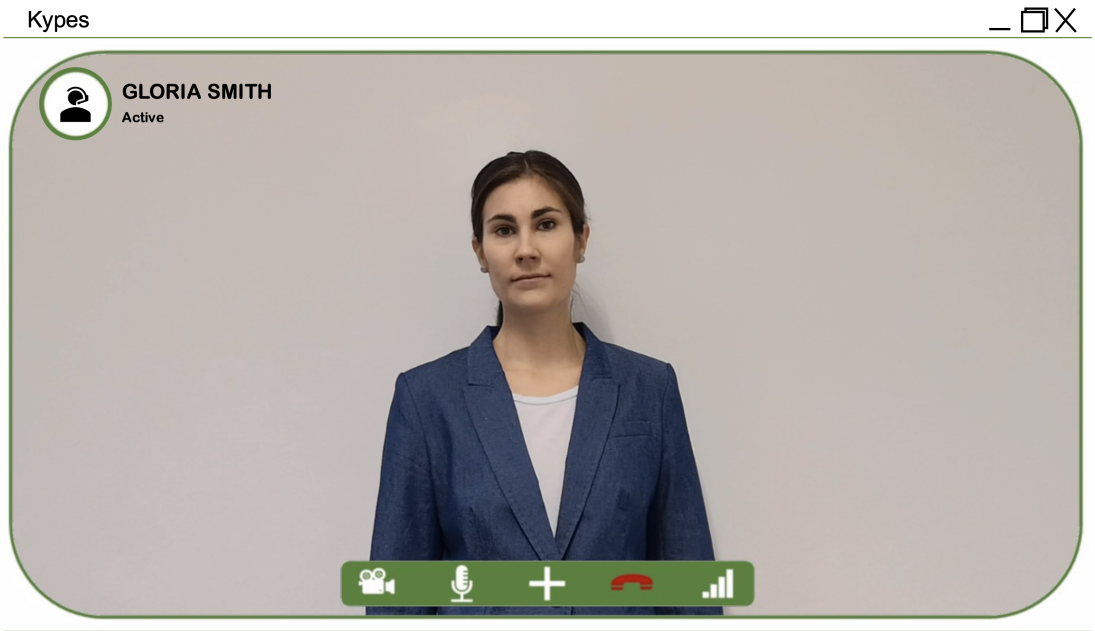


Figure 6.2: The human instructor.

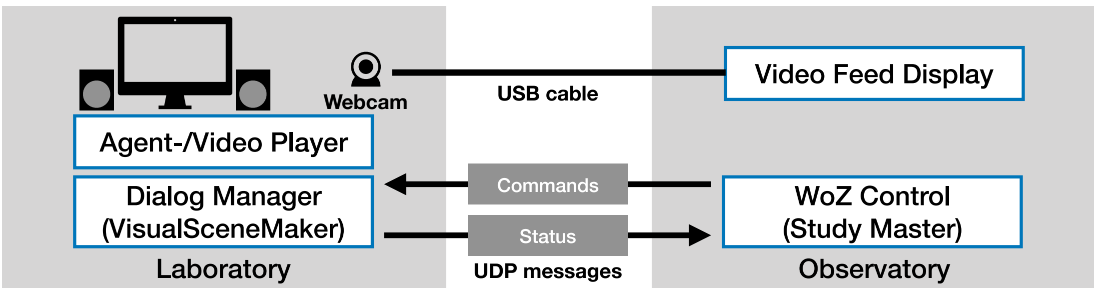


Figure 6.3: System setup and components.

6.3.5.2 VisualSceneMaker and StudyMaster

The interaction for both instructors was scripted with the VisualSceneMaker (VSM; Gebhard et al., 2012), a real-time execution and authoring tool for modeling verbal and non-verbal behavior of virtual agents. The logic behind the play (e.g., which scene to play when, or what process to run in the background) is determined by a finite-state automaton called the *scene flow*. What happens in the scene (the text for each character, animations, etc.) is described in the *scene script*, which is a text file. In our experiment, we created a project for each condition and VSM ran those on the computer in the lab. In the virtual agent condition, VSM immediately controlled the behavior of the virtual agent. In the human instructor condition, VSM was used to play the pre-recorded videos with the VLC Player.

The *StudyMaster* is a tool to remote control the VSM by sending (User

Datagram Protocol) network messages. These messages contain information on how to change variables in the scene flow and thereby influencing it. The VSM itself sends messages back containing status information. The experimenter used the StudyMaster running on a tablet to control the behavior of the instructor.

6.3.6 Measurements

Assessment of the instructor was measured with one self-constructed item respectively on a 5-point scale from 1 (*disagree strongly*) to 5 (*agree strongly*) for attractiveness, sympathy and authority. Trust in the instructor was measured with five own items on a 5-point scale. Items were “I trust Gloria”, “Gloria seems sincere to me.”, “I think Gloria means well to me.”, “I fear that Gloria wants to harm me.”, “I feel uncomfortable in the presence of Gloria.” (Cronbach’s Alpha .80).
Stress was measured before and after the tasks with the short version of the State-Trait-Stress-Inventory (Marteau & Bekker, 1992) translated in German. The STAI-6 raises the acutely felt stress with six items on a 4-point scale ranging from 1 (*not at all*) to 4 (*very*). Cronbach’s Alpha was .79 for the pre-test and .84 for the post-test.
Shame was measured before and after the tasks with six shame items from referring scales of the German versions of the Differential Emotion Scale (DES) (Izard et al., 1993; Merten & Krause, 1993) and the Positive And Negative Affect Schedule (PANAS) (Krohne et al., 1996; Watson et al., 1988). Two own items (“indignant” and “abashed”) were added. To avoid priming, especially before the tasks, we included 34 other items of the DES as well as the PANAS. Items had to be answered on a scale ranging from 1 (*not at all*) to 5 (*extremely*). Cronbach’s Alpha was .90 for the pre-test and .93 for the post-test.
General Trust was measured on a 5-point scale with six items (Kramer, 1999), e.g., “Most people are trustworthy.”. Cronbach’s Alpha was .66.
Personality was measured with the German version of the Big Five Inventory (BFI) (John et al., 1991). For the self-assessment of openness, conscientiousness, neuroticism, agreeableness, and extraversion, 42 items had to be rated on a 5-point scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Cronbach’s Alpha for the five scales was between .78 and .90.

6.4 Results

In general, 45% of the subjects fulfilled all 18 tasks. The most frequent breaking task was throwing down a coffee cup (17%, task 15), followed by telling a joke (15%, task 6) and taking out a booger from the nose (13%, task 17). On average, 14.35 ($SD = 5.00$) tasks were performed.

6.4.1 Assessment of the Instructor

The human instructor ($M_{\mathrm{Att}} = 3.67$, $SD_{\mathrm{Att}} = 0.88$; $M_{\mathrm{Sym}} = 3.50$, $SD_{\mathrm{Sym}} = 0.90$) was assessed more attractive ($F(1{,}56) = 21.56$, $p < .001$, $\eta_p^2 = .278$) and more sympathetic ($F(1{,}56) = 5.17$, $p < .05$, $\eta_p^2 = .085$) than the virtual agent ($M_{\mathrm{Att}} = 2.63$, $SD_{\mathrm{Att}} = 0.85$; $M_{\mathrm{Sym}} = 2.93$, $SD_{\mathrm{Sym}} = 1.05$). Regarding their perceived authority the human instructor did not differ ($F(1{,}56) = 0.36$, $p = .48$) from the virtual agent.

Moreover, the trust in the instructor did not differ between the groups ($F(1{,}58) = 2.90$, $p = .09$). However, trust in the instructor correlated significantly with the perceived shame ($r = .32$, $p < .001$) and stress ($r = .30$, $p < .05$) of the participants.

6.4.2 Hypotheses

Hypotheses 1 and 2b tested for a non-existent difference. Therefore, the classic statistical tests, a t-test and a multivariate analysis of variance, is enriched with the Bayes Factor, that allows researchers to express preference for either the null hypothesis or the alternative (Rouder et al., 2009).
Hypothesis 1 stated that obedience, measured with the breaking task, does not differ between the human and agent instructor. We found no significant difference for the breaking task between the conditions ($t(58) = -0.13$, $p = .90$). The Bayes Factor was in favor for the null hypothesis (*JSZ-B_{01}* = 5.11, *Scaled-Information-B_{01}* = 3.97). With the virtual agent, participants finished on average 14.27 (*SD* = 4.90) tasks and with the human instructor 14.43 (*SD* = 5.18). Hence, hypothesis 1 was supported by our data.
Hypothesis 2a proposed that obedience to stressful and shameful tasks leads to a higher self-reported stress and shame level. Our data (Table 6.2 for descriptive data) showed that after the experiment the self-reported stress and shame values are significantly higher than before ($F_{\mathrm{Stress}}(1{,}59) = 12.33$, $p < .001$, $\eta_p^2 = .173$; $F_{\mathrm{Shame}}(1{,}59) = 60.49$, $p < .001$, $\eta_p^2 = .506$). Therefore, hypothesis 2a was supported by our data.
Hypothesis 2b posited that the level of stress and shame after the experiment does not differ between the groups. Adding condition in the multivariate model used for 2a, we could not find a significant difference between the agent and the human instructor (*Wilks*-$\lambda = .978$, $F(2{,}57) = 0.63$, $p = .534$). Neither the single t-tests, needed for the Bayes Factor, did show a significant difference between the conditions for stress ($t(58) = -0.70$, $p = .487$) or shame ($t(58) = -0.34$, $p = .739$). The Bayes Factor showed a preference for the null hypothesis for stress (*JSZ-B_{01}* = 4.12, *Scaled-Information-B_{01}* = 3.17) as well as for shame (*JSZ-B_{01}* = 4.89, *Scaled-Information-B_{01}* = 3.79). Overall, hypothesis 2b that the agent invoked the same level of stress and shame like the human instructor was supported by our data.

Table 6.2: Descriptives for self-reported shame and stress before and after the task fulfillment.

	Agent *M* (*SD*)	Human *M* (*SD*)	Overall *M* (*SD*)
$\text{Stress}_{\text{pre}}$	1.86 (0.47)	2.03 (0.53)	1.94 (0.50)
$\text{Stress}_{\text{post}}$	2.23 (0.69)	2.23 (0.58)	2.23 (0.63)
$\text{Shame}_{\text{pre}}$	1.43 (0.62)	1.61 (0.69)	1.52 (0.66)
$\text{Shame}_{\text{post}}$	2.47 (1.11)	2.42 (0.96)	2.44 (1.03)

Note. $N = 60$. Stress was measured on a 4-point scale, shame was measured on a 5-point scale. Pre stands for the self-reported values before the task fulfillment, post for the self-reported values after the task fulfillment.

Hypothesis 3 stated that there is an interdependency between obedience and personality. The correlations of the five personality factors with obedience, measured with the breaking task, did not reach significant levels. The hierarchical linear regression with the five factors ordered descending according to the strength of the correlation (Field, 2013), did not reach significance.

6.5 Discussion

This study aimed to compare a virtual agent giving instructions to fulfill stressful and shameful tasks with a human instructor. Both instructors were presented in a video-chat under the cover-story of a creativity test. Participants had to obey to a maximum amount of 18 tasks increasing in difficulty. The difficulty level was empirically justified in a pre-study. Our results show that participants obey the virtual agent at the same level as the human instructor. On average, around 14 tasks were fulfilled in both conditions. Moreover, we found that obedience in fulfilling shameful tasks increases the level of stress and shame. This increase was independent from the instructor (human vs. virtual agent). The virtual agent and the human invoke feelings of stress and shame to the same amount. Additionally, we examined the influence of personality on obedience to shameful tasks, but could not find any effects.

Our finding that participants obey towards non-humans like towards human instructors is consistent with previous work examining robots as instructors (Cormier et al., 2013; Geiskkovitch et al., 2016; Menne, 2017).

Likewise the classic and more recently adapted obedience experiments (Milgram, 1965; Slater et al., 2006), we found that obedience in fulfilling shameful tasks influences participants' self-reported stress and shame. The level of stress and shame increased significantly from before to after the task fulfillment.

Several studies conclude that with a higher degree of anthropomorphism people feel shame also towards non-human entities like virtual characters or robots

(Bartneck et al., 2010; Kang & Gratch, 2010; Lucas et al., 2014). In our study, we could find similar results. Participants reported shame in the post-questionnaire when instructed by a virtual agent, although shame is an interpersonal emotion, i.e., it occurs typically or almost only in the presence of an emotional relation to another human (Izard, 1977). Even more, the virtual agent does not only invoke shame, but it also invokes the same amount of shame and stress like the human instructor. This goes in accordance with findings by Menne, who showed that non-human entities are able to invoke the same feelings of shame like humans (Menne, 2017).

Regarding the correlation between personality and obedience to shameful tasks, our data did not show significant correlations of the five factors openness, conscientiousness, neuroticism, agreeableness, and extraversion with the amount of fulfilled tasks. The hierarchical linear regression did not show any significant predictions when ordering the factors descending according to the strength of the correlation. Those findings are in line with previous studies (Burger, 2009; Milgram, 1974), but not with (Bègue et al., 2015), where a positive correlation between agreeableness and obedience was found.

6.5.1 Limitations and Future Work

Likewise the virtual agent, the human instructor was presented in a video-chat setting. Compared to other studies that had real humans in face-to-face interactions, a human in a video-chat might not have the same authority. A simple reason for this might be that it is not physically present and therefore a smaller threat. However, we could show that almost half of the participants fulfilled all tasks even from an entity that is not physically present. Therefore, our study indicates that the physical presence of the instructor might not be that important as Milgram (1965) stressed out. Moreover, guaranteeing a standardized behavior of a human instructor seeing 30 people performing, for example, the chicken dance, is nearly impossible.

For the assessment of stress and shame, we relied on self-reported data of the participants. Our study design did not include objective measures of stress like in (Slater et al., 2006). Also, shame was not measured objectively, for example, by analyzing social signals of shame or shame regulation (Gebhard et al., 2018). Therefore, future work should include the objective analysis of social signals of shame and shame regulation. This observational data can be an appropriate method to validate self-reported shame and stress.

In our experiment, participants had to fulfill shameful tasks. Therefore, we do not know if the increase in the stress and shame levels in both conditions is due to the obedience or due to the execution of the tasks. Future work could consider using more neutral tasks to find out if obedience itself leads to an increase in the stress and shame levels.

This work compares a human and a virtual agent in a video-chat as instructors. Future research could include other instructors like a present robot or a robot in a video chat. Moreover, to examine the influence of anthropomorphism the virtual agent could be compared to a conversational agent without representation.

6.6 Conclusion

The results of this study indicate that humans obey virtual agents just as they do towards humans in a video-chat. We could show that participants fulfill the same amount of shameful tasks independent of the instructor. Moreover, both instructors were able to elicit the same level of shame and stress in the participant. Therefore, our results provide one more indication for the validity of the Media Equation. Virtual agents seem to be able to influence humans even when it comes to tasks that are uneasy to perform.

6.7 Acknowledgment

We thank Charamel GmbH for realizing our requirements with regard to the virtual agent and Saarland Informatics Campus for the provision of the experiment rooms.

7

Developing a Social Biofeedback Training System for Stress Management Training

Tanja Schneeberger, Naomi Sauerwein, Manuel S. Anglet, and Patrick Gebhard

German Research Center for Artificial Intelligence

Published in Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)
https://doi.org/10.1145/3397481.3450683

This chaper contains the original version (with slight changes due to formatting) of our full paper presented at the 26th International Conference on Intelligent User Interfaces (IUI '21). During the online conference, the first author presented the paper orally to the conference attendees. The paper received a honorable mention as outstanding paper.

Abstract
Coping with stress is critical to mental health. Prolonged mental stress is the psychological and physiological response to a high frequency of or continuous stressors, which has a negative impact on health. This paper presents a virtual stress management training using biofeedback derived from the cardiovascular response of the heart rate variability (HRV) with an interactive social agent as biofeedback trainer. The evaluation includes both, a subject-matter expert interview and an experiment with 71 participants. In the experiment, we compared our novel stress management training to a stress management training using stress diaries. The results indicate that our social agent-based stress management training using biofeedback significantly decreased the self-assessed stress levels immediately after the training, as well as in a socially stressful task. Moreover, we found a significant correlation between stress level and the assessment of one's performance in a socially stressful task. Participants that received our training assessed their performance higher than participants getting stress diaries. Taken this together, our novel virtual stress management training with an interactive social agent as a trainer can be evaluated as a valid method for learning techniques on how to cope with stressful situations.

Keywords: Social Agents; Stress Management Training; Biofeedback; Mental Health

7.1 Introduction

Persistent mental stress is a massive problem in today's society. Repeated, excessive, or prolonged stress reactivity can increase health risks. Many mental and physical diseases are caused by persistent mental stress (Cohen et al., 2007; DeLongis et al., 1988), like major depression and depressive symptoms (Hammen, 2005; Mazure, 1998; Monroe & Simons, 1991) or cardiovascular disease (Kivimäki et al., 2006; Krantz & McCeney, 2002; Rozanski et al., 1999). The World Health Organisation considers stress to be one of the most significant health risks of our time and expects that by the end of 2020 more than one in two sick days will be caused by stress (WHO, 2001).

Stress management training, in general, was found to have positive effects, for example, on a healthy population (Chiesa & Serretti, 2009), in the work context (Richardson & Rothstein, 2008), or on students (Kraag et al., 2006). One method of stress management training is biofeedback training, in which immediate and continuous feedback for variations in physiological activity is provided to the trainee. With the guidance of a biofeedback coach (often a therapist), the goal is to raise awareness for usually unconscious physiological functions and to reach voluntary control of them (Schwartz & Andrasik, 2017).

With the advanced simulation of social skills (e.g., active listening, mimicry,

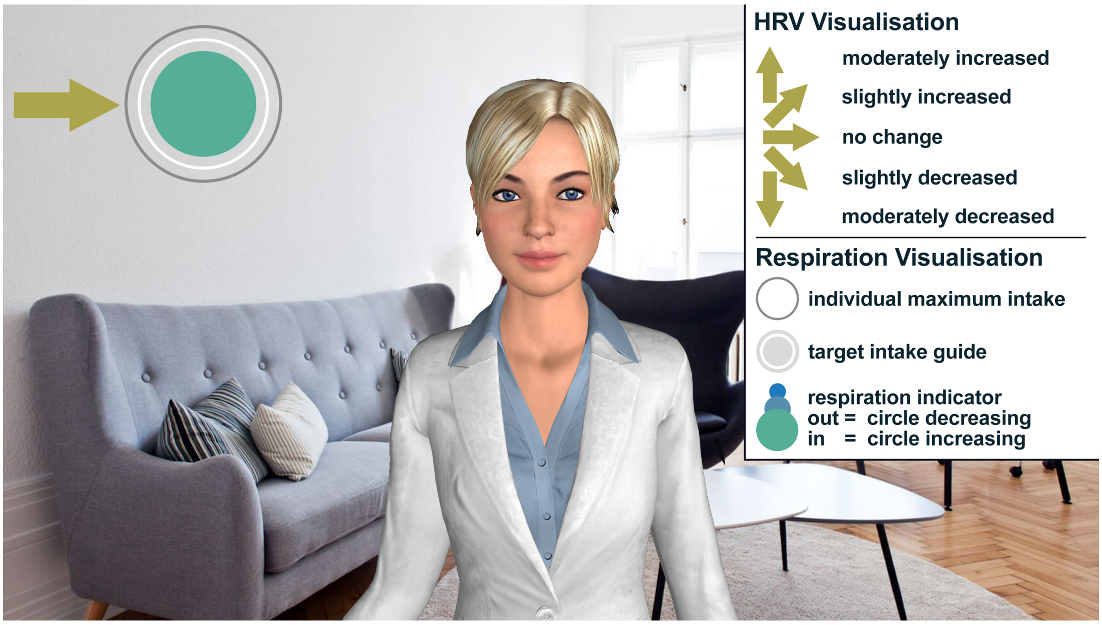


Figure 7.1: Biofeedback training user interface with the interactive social agent Gloria.

gestures, emotion models) of interactive agents, these are becoming able to emulate social relationships and are employed as virtual coaches (Bickmore & Picard, 2005; Gebhard et al., 2018; Gratch et al., 2013; Gratch et al., 2007). Since emulating social relationships is essential to tackle mental grievances (Davis & Hadiks, 1994) like stress, such agents could, in principle, be used to improve technologically supported stress-management by emulating the interaction and tasks of a human biofeedback coach.

This paper presents a technological improved version of the virtual stress-management training with the interactive social agent Gloria (Schneeberger et al., 2020) (Figure 7.1). Moreover, it presents the results of an experiment with 71 participants assigned to either the experimental or control group aiming to raise awareness for often unconscious patterns.

The biofeedback training for the experimental group consists of two phases: 1) Visual feedback and Gloria; and 2) Gloria only. In the first one, both a biofeedback monitor, displaying information about physiological functions and Gloria are presented. Gloria gives background information and feedback on proper sensor handling. In the second one, the biofeedback monitor is removed, and Gloria's guidance replaces its information. She verbally gives encouraging feedback if the physiological functions develop in a non-optimal direction. The purpose of the two-phase training was to fade out the biofeedback monitor as technically-supported awareness tool to focus more on the social situation between the trainee and the coach. The presented approach's unique feature is that the biofeedback training instructions are given entirely by a social agent instead of a human coach.

The diary-based stress management training for the control group focuses on the conscious reflection about stressful life events to raise awareness for these and establishes order and overview for stressful situations.

In the experiment, the novel biofeedback training approach is compared to the widely used diary-based stress management training. To evaluate the transfer of the trained stress-management strategies, we measured the experienced stress of the study participants in a human-human social situation without any technical support.

7.2 Background on Stress and Biofeedback Training

A stress reaction is an unspecific reaction of a living being to a stimulus from its environment and at first value-neutral (Selye, 1946, 2013). What stimulus and intensity it has to arise to be experienced as negatively stressful is highly subjective. According to the transactional model of stress (Lazarus & Folkman, 1984), it depends on the perception, interpretation, and evaluation of a person's environment. A stimulus, therefore, only becomes a negative stressor when it is evaluated as such (Lazarus, 1993). This evaluation and possible personal stress intensifiers lead to a stress reaction, a physical as well as a psychological deviation of the homeostasis. In other words, "Stress arises when individuals perceive that they cannot adequately cope with the demands being made on them or with threats to their well-being" (Lazarus, 1966). In response to a stressor, the sympathetic nervous system is activated. In addition to the release of the neurotransmitters, adrenalin and noradrenalin, the sympathetic nervous system's impulses primarily cause physiological reactions to prepare the body for a quick response to the stressful situation (De Kloet et al., 2005). This includes an increase in heart rate (Taelman et al., 2009), respiratory rate (Masaoka & Homma, 1997), and a decrease in heart rate variability (HRV) (Taelman et al., 2009). A high frequency or continuity of stressors results in prolonged mental stress that has a negative long-term impact on psychological and physiological health. One reason for that is that the body cannot reach its homeostasis (Cohen et al., 2007), which can cause psychiatric disorders such as depression (Bao et al., 2008). As a physiological reaction, prolonged mental stress significantly reduces HRV (Castaldo et al., 2015). Different coping strategies to adapt or cope with stress can be used to respond to stress reactions (Carver et al., 1989; Lazarus, 1966) and can be learned in stress management trainings.

One instrument of stress management training is biofeedback training. Biofeedback training is based on the process of becoming aware of unconscious body's physiological processes using specialized devices and sensors (Brown, 1977). It is a widely used method for teaching voluntary control of various physiological functions by providing immediate and continuous feedback for physiological activity variations (Schwartz & Andrasik, 2017). Mostly, it requires a trainer to be present for the entire training (Mikosch et al., 2010; Munafo et al., 2016; Paul & Garg, 2012; Prinsloo et al., 2013; Prinsloo et al., 2011; Sherlin et al., 2009). Feedback is usually given in the form of visual and/or auditory signals derived from physiological recording devices. Physiological functions chosen for biofeedback training can include muscle tension, finger temperature, heart rate, blood pressure, or HRV (Lehrer, 2007).

HRV represents the beat to beat changes in the inter-beat interval (time between two successive R-waves). HRV training aims to increase the HRV amplitude that promotes vegetative nervous system balance. This balance is associated with improved physiological functioning as well as psychological benefits. HRV biofeedback appears to have profound effects across systems (Lehrer, 2013) and has been meta-analytically evaluated to significantly reduce self-reported stress and anxiety for both community and clinical settings (Goessl et al., 2017). It has been successfully applied in healthy populations in recent years, especially in highly stressful work environments (Ratanasiripong et al., 2015; Sutarto et al., 2010). Factors that influence heart rate and HRV are respiratory depth and interval (Eckberg, 1983). Slow breathing rates cause an increase in heart rate variability. Therefore, most individuals can learn to quickly increase their HRV amplitude by slowing the breathing rate (around six breaths/min) to each individual's resonant frequency at which the amplitude of HRV is maximized (Bernardi et al., 2001; Sutarto et al., 2010).

In this paper, an HRV-based biofeedback training is presented. It uses a three-channel (two sensors just below the left and right collarbone, one at the abdomen's center, Figure 7.2 User with Sensors) to record the heart rate as an output signal for heart rate variability. As an initial influencing process, the respiratory rate is recorded with a sensor-equipped chest strap and feed-backed to the user.

7.3 Related Work

The health context, as a use case for interactive social agents, has been getting attention in research for about 15 years.

The Fit Track with the relational agent Laura is one of the first systems in this area (Bickmore & Picard, 2005). Laura has the role of an exercise advisor that interacts with patients daily for one month to motivate them to exercise more. Laura was equipped with different communication skills (i.a., empathy, social dialogue, nonverbal immediacy behaviors) to build and maintain good working relationships over multiple interactions. A study showed that the use of those social behaviors significantly increases the working alliance and the desire to continue working with the system.

Lucas et al. (2014) showed positive effects of autonomous social agents in a health-care setting on overcoming the barrier to receive truthful patient information. They compared two different interviewers in a health-screening interview. Participants interacted with a virtual human and were led to believe that either a human or automation controlled the virtual human. Participants who believed they were interacting with a computer, reported lower fear of self-disclosure and were rated by observers as more willing to disclose.

The potential use of virtual humans as counselors in psychotherapeutic situations was investigated by Kang and Gratch (2010). Examining self-disclosure of patients in psychotherapy, they analyzed with which conversational partner participants disclose more private information measured with self as well as with external assessment. Their study reveals that a virtual human can elicit more self-disclosure in a hypothetical conversational scenario than a human in a raw or degraded video.

A system for PTSD patients (Tielman et al., 2017), in which patients can recollect their memories in a digital diary and recreate them in a 3D WorldBuilder, is using a virtual agent to inform and guide patients through the sessions. The agent employs an ontology-based question module for recollecting traumatic memories to elicit a detailed memory recollection further. In their study, the authors found hints that these questions were useful for memory recollection and conclude that their system can be a valuable addition in PTSD treatments, offering a novel type of home therapy.

Zhang et al. developed a virtual conversational agent that provides cardiovascular health counseling to hospitalized geriatrics patients (Zhang et al., 2015). The agent counsels patients on several health-related aspects such as decreasing stress and motivating them to be more involved and proactive in their self-care.

An embodied conversational agent in the role of a virtual intelligent university student advisor (Kavakli et al., 2012) was realized and piloted to support undergraduate students in stress management during their exams. In their study, the authors focus on gender effects of the advisor's perceived pleasantness, credibility, clarity, dynamism, and competence. Voices of male advisors were assessed as more pleasant and credible than female advisors', voices of female advisors were considered as more clear, dynamic and competent.

Shamekhi et al. (2016) developed a virtual coach system for patients with a spinal cord injury that need training and support for self-care management after hospital discharge. The virtual coach educated the patients about managing their health and motivated them to healthy behavior. In their exploratory study, the authors found that patients were highly receptive and evaluated the virtual coach as an effective medium to promote self-care.

There are also systems in which biofeedback training is technologically supported. Chittaro and Sioni (2014) employ virtual agents to reflect the user's level of stress. Based on the detection of the user's current stress level, the virtual agent's affective state and behavior are adapted as a form of embodied feedback. This work focuses on a comparison of single and multi-sensor stress detection algorithms. It uses the embodied feedback given by the virtual agent as a mediator for the perceived quality (in terms of accurateness) of biofeedback.

A virtual reality-assisted HRV biofeedback for golfers was introduced by Lagos et al. (2011). In a 10 weeks HRV biofeedback training, golfer and coach met at a virtual reality golf center to practice skills for breathing at resonance frequency during golf performance. In a case study with one golfer, after the training, reduction in symptoms of anxiety, stress, and sensation seeking and increases in total HRV, and sport performance improvement was observed.

Generally, it seems that the exploitation of social agents in health-care might have unique possibilities and even advantages to support the health-care system to overcome existing challenges. Moreover, it appears that biofeedback training can profit from a technological enhancement (Rockstroh et al., 2019). However, none of the current systems provides an interactive social agent that is giving biofeedback training instructions.

7.4 Realization

The interactive biofeedback environment extends the existing system (Schneeberger et al., 2019a) with biosignal interpretation, biofeedback monitor, and image display components. The environment features the interactive social agent Gloria as the main user interface. Gloria gives biofeedback training instructions and comments, as a human biofeedback coach would do.

Hardware-wise, the system runs on a PC running MS Windows 10™ (Intel Core i7 CPU@3.5GHZ, 16GB Memory, NVIDIA RTX 2080 graphics cards) connected to a computer monitor (40 inches), showing Gloria at a realistic size. The interaction with the system is realized with wireless biosignal sensors that measure the physiological parameters respiration rate and heart rate. To measure the physiological parameters, we rely on the *Plux* wireless biosignal toolkit[1]. These sensors were explicitly developed for research purposes and have already been successfully used in various studies (Bosse et al., 2016; Islam et al., 2014; Schek et al., 2017). To derive the heart rate, we used an electrocardiogram local differential triode. The three electrodes were placed below both participants' collarbones and centrally under the costal arch. To derive the respiratory rate and depth, we used a piezoelectric respiration sensor-equipped chest strap that was placed over the clothing. Both sensors were connected to a wireless 4-channel hub. To reduce data noise, the hub was placed on the table in front of the participant and all participants were instructed to sit as still as possible and to switch off Bluetooth

[1] plux.info

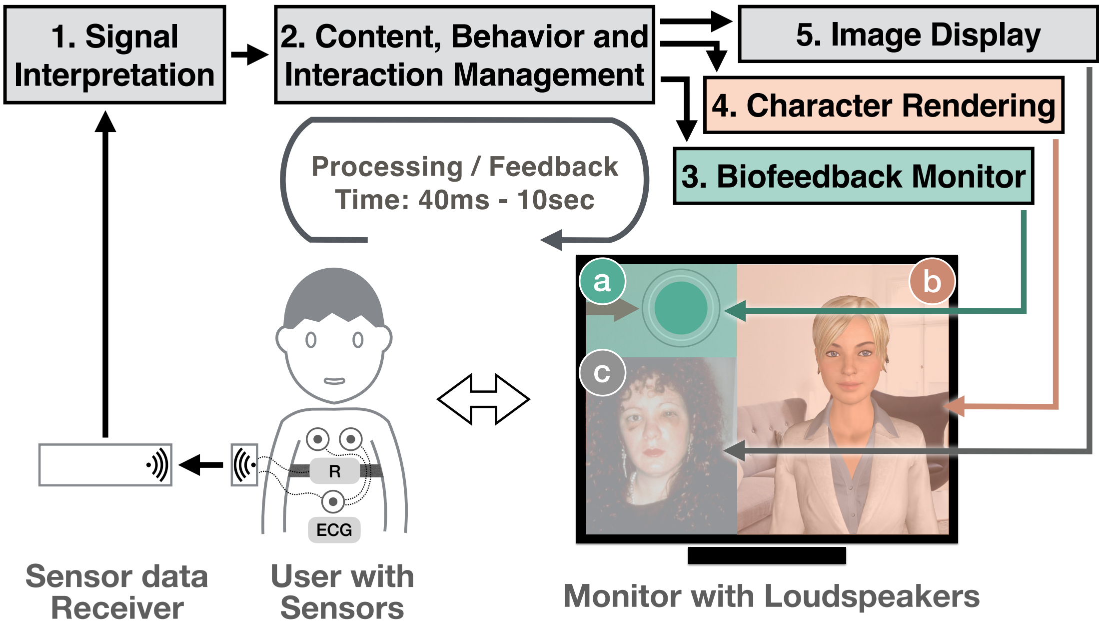


Figure 7.2: System architecture.

on their devices.

Software-wise, the interactive biofeedback environment (Figure 7.2) is realized with five main software components: 1. real-time signal interpretation, 2. content, behavior and interaction management, 3. biofeedback monitor, 4. character rendering, and 5. image display. All components are implemented as software agents and are asynchronously coordinated with events exchanged by a UDP network architecture. The last three components produce graphical output that is displayed in separate sections on the computer monitor (Figure 7.2, a - c).

Signal interpretation To process raw electrocardiogram data we employ the open-source *Social Signal Interpretation* framework (*SSI*, Wagner et al., 2013). SSI has a pipeline concept that allows parallel processing of multiple sensor streams in real-time. In an electrocardiogram, the HRV describes the difference between two successive heartbeats. Thus, to assess the HRV, we used the Root Mean Square of the Successive Differences (RMSSD), one of the most commonly used measures derived from interval differences (Malik, 1996). RMSSD is particularly suitable in the short-term range and permits a reliable analysis of the HRV (Nussinovitch et al., 2011). For the visual and verbal feedback of the training, the RMSSD values from ten heartbeats were calculated. For this purpose, absolute timestamps were assigned to the ten data points, which were subsequently sorted to exclude first-in-first-out errors. To avoid sorting errors, two additional values were added as buffers before and after, which were not included in the HRV calculation.

Content, behavior and interaction management Central to the realization is the open-source *VisualSceneMaker* (VSM) toolkit (Gebhard et al., 2012), which is used to coordinate all software agents. VSM comes with a real-time execution and authoring component for modeling verbal and non-verbal behavior of virtual agents as well as system actions. Their execution (e.g., which scene to play when, or what process to run in the background) is determined by a finite-state automaton called the *sceneflow*. What happens in the scene (the agent's utterances or animations, or system commands) is described with a *scene script*, which is a human-readable text file.

Biofeedback monitor To display the HRV and the respiratory rate and depth, we realized a biofeedback monitor in *Unity*[2]. A yellow colored arrow visualizes the HRV. Every ten seconds, it gives feedback in five steps (Figure 7.1 up right) on how the HRV developed regarding the previous value. The respiration is visualized with three circles (Figure 7.1 up left): 1. the grey one indicates the individual adaptive maximum intake that defines the deepest inhalation as a new maximum 2. the white pacing stimulus that moves continuously from the center to the grey circle and functions as the target intake guide (three seconds inhalation, four seconds exhalation), 3. the blue to green one that represents the actual participant's respiration (approximately 40ms delay).

Character rendering Gloria is a high-quality agent with a natural human appearance and verbal as well as nonverbal dialogue skills. Gloria is capable of performing social cue-based interaction with the user. She performs lip-sync speech output using the state-of-the-art CereProc[3] Text-To-Speech system. For a more advanced animation control, Gloria allows the direct manipulation of skeleton model joints (e.g., the neck joint or the spine joint). She comes with 36 conversational motion-captured gestures, which can be modified during run-time in some aspects (e.g., overall speed, extension). Besides, the agent comes with a catalog of 14 facial expressions, which contains, among others, the six basic emotional expressions defined by Ekman (1992).
Gloria is rendered by the commercial Charamel rendering engine that is free to use for any research purposes[4].

Image display During the training sessions, stress-inducing pictures and Stroop tables (cf. Figure 7.3) were displayed.

[2]unity3d.com
[3]cereproc.com
[4]charamel.com

7.5 Evaluation Outline

We evaluated the biofeedback trainer with a mixed-methods design applying both qualitative and quantitative approaches. In an interview with a biofeedback trainer, we gathered qualitative data giving us insights about the potential of the virtual biofeedback training with the social agent. In a user study, we gathered quantitative data comparing the virtual biofeedback training against a control group regarding its effectiveness in teaching stress management in stressful situations. As planned, the data was collected from December 2019 until the beginning of February 2020 (without the influence of the pandemic situation in 2020).

7.5.1 Subject-Matter Expert Interview

After implementing the system, we conducted a subject-matter expert (SME) interview with a biofeedback coach. The interview started with a presentation and explanation of the system. Afterward, there was an open discussion regarding the usefulness, potential, and applicability. The interview was conducted before the start of the user study, in case the SME recommended major adaptations. However, there were no adaptations requested.

7.5.2 User Study

To evaluate the realized biofeedback training, we designed a two-session intervention with the biofeedback training for the experimental group (EG). For the control group (CG), a two-session online training using stress diaries took place during this period. Both training approaches aim to gain consciousness on often unconscious patterns. After the training, we used a stress-inducing task to examine participants' stress regulation. We expected that the virtual biofeedback training positively influences the subjective stress perception during stressful tasks due to a more conscious regulation of the heart rate variability (HRV) in stressful situations. Thus the virtual biofeedback training for the experimental group should have a relaxing effect. Furthermore, the positive impact of higher HRV on the cognitive and emotional levels was expected to influence subjective performance assessment positively. For the control group participants who keep mindfulness-promoting stress diaries, we assumed weaker intervention effects.
Hypothesis 1a and 1b: The virtual biofeedback training leads to a higher general relaxation compared to the mindfulness-promoting stress diaries. After the training intervention, the general level of stress is dependent on the intervention. *1a*: Participants in the EG report a lower stress level, measured with a stress questionnaire, after the training before the stress-inducing task than participants in the CG ($\text{Stress}_{\text{pre}}\text{EG} < \text{Stress}_{\text{pre}}\text{CG}$). *1b*: Participants in the EG report a

lower current stress level, measured with a visual analog scale, after the training before the stress-inducing task than participants in the CG (Current $\text{Stress}_{\text{pre}}$EG < Current $\text{Stress}_{\text{pre}}$CG).

Hypothesis 2a, 2b and 2c: The virtual biofeedback training leads to a better stress regulation in the stress-inducing tasks. *2a*: Participants in the EG indicate a lower stress level, measured with a stress questionnaire, after the stress-inducing task than participants in the CG ($\text{Stress}_{\text{post}}$EG < $\text{Stress}_{\text{post}}$CG). *2b*: Participants in the EG indicate a lower current stress level, measured with a visual analog scale, after the stress-inducing task than participants in the CG (Current $\text{Stress}_{\text{post}}$EG < Current $\text{Stress}_{\text{post}}$CG). *2c*: Participants in the EG assess the stress-inducing task, measured with a visual analog scale, as less stressful than participants in the CG (Task stressfulness EG < Task stressfulness CG).

Hypothesis 3a, 3b and 3c: Higher HRV has a positive effect on the cognitive (A. L. Hansen et al., 2004) and emotional levels (Geisler et al., 2013), which should positively influence the subjective performance assessment. *3a*: Stress level and the assessment of one's own performance are dependent. The higher the stress level after the stress-inducing task, the lower the self-assessed performance. *3b*: Task stressfulness and the assessment of one's own performance are dependent. The more stressfulness the task is experienced, the lower the self-assessed performance. *3c*: Participants in the EG assess their performance during the stress-inducing task better than participants in the control group (Self-performance EG > Self-performance CG).

7.6 Methods

7.6.1 Participants

In total, 71 participants (41 female, 30 male) participated in the experiment. All participants were randomly assigned to the experimental condition ($n_{\text{EG}} = 35$) or the control condition ($n_{\text{CG}} = 36$). They were recruited mainly via social network groups of psychology students on condition that they were fluent in German, between 17 and 65 years and without current mental illness. Psychology students were rewarded with course credit for participation, students from other faculties and employed people were rewarded with hot drink vouchers for coffee shops on campus. Participants were aged between 17 and 61 years ($M = 28$ years, $SD = 8.8$ years). There was no significant difference ($t(69) = 0.91$, $p = .30$, $d = .22$) regarding the general stress level measured with the Perceived Stress Scale between the experimental group ($M = 2.82$, $SD = 0.62$) and the control group ($M = 2.97$, $SD = 0.73$) before starting the experiment.

7.6.2 Procedure

Experimental Group. Participants in the experimental group had two face-to-face sessions with one recreation day between the two sessions (Figure 7.3). After welcoming the participant to the first session, the experimenter explained the course, duration, and goal of the session. The participants were given introductions into stress, its consequences, and biofeedback training techniques, especially heart rate variability. After filling out the informed consent form, the demographic questionnaire, and the perceived stress questionnaire, they were equipped with the bio-signal sensors and led to the biofeedback training. After sitting down at a table in front of the display at a distance of 85 cm, the experimenter handed over to the virtual coach and left the room. After the 45 min training session, the experimenter entered the room, was handed the bio-signal sensors back, and released the participant. Two days later, the second session started similarly and was followed by a 35 minutes training session. After that, participants filled in the stress questionnaire, a visual analog scale (VAS) regarding their current stress perception, as well as the user experience questionnaire. Then, after five minutes break, the stress test started (Sec. 7.6.3.4). Afterward, they were given the stress questionnaire and three visual analog scales (current stress, task stressfulness, and self-assessed performance). In the end, the participants were debriefed and released. In total, the experimental group was involved for 140 minutes.

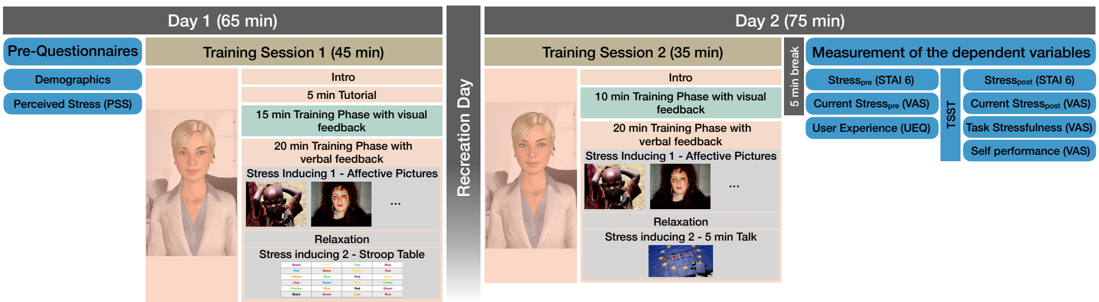


Figure 7.3: Procedure for the experimental group with measurements in rounded boxes.

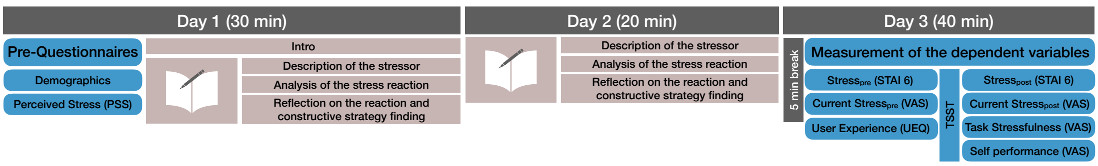


Figure 7.4: Procedure for the control group with measurements in rounded boxes.

Control Group. We used an active control group that got a diary-based stress management training (Figure 7.4). Writing about personally experienced stressors has been associated with improvements in mental and physical health in numerous investigations (Ullrich & Lutgendorf, 2002) and was shown as a valid method to cope with stress (Alford et al., 2005; Donnelly & Murray, 1991). Stress journaling is a widely used method in cognitive behavioral therapy and is often given as homework for the patient. Compared to the biofeedback training, it represents more the state-of-the-art therapists' possibility to train stress management. Participants in the control group, filled in a stress diary that they got via e-mail on two consecutive evenings. On the third day, they had a face-to-face session with the experimenter to measure the dependent variables. After arrival, the participants could calm down for five minutes before the measurement of the dependent variables started. The procedure was similar to the experimental group. In total, the control group was involved for 90 minutes.

7.6.3 Material and Set-up

7.6.3.1 Procedure of the Biofeedback Training

The biofeedback training consisted of two training sessions that participants fulfilled on two days with one recreation day between the two sessions.

At any point in time, Gloria is aware of any problems concerning the biofeedback sensors and data processing. If due to the participant's extraordinary movement, the sensor data is flawed because of too many noise or false ECG data, Gloria explains the situation. She asks the user to find a comfortable position in which s/he can sit relaxed. After this explanation, she resumes the training. Repetitive occurrences of such situations are covered with different variations and suggestions (5 variations total) to keep the illusion that Gloria comes with some level of (social) competence.

After welcoming the participant for the first training session, the virtual coach Gloria introduced herself and the topic of stress. This was followed by a five-minute baseline measurement in which the participants were asked to relax and close their eyes in order to reach a relaxed basic level, regardless of the situation from which they had just come to the study. During the next phase, the tutorial phase, participants were introduced to the biofeedback technique, the breathing circle (Figure 7.1 up left) was explained, and a five-minute testing phase was offered. Then the arrow for visual feedback was introduced (Figure 7.1 up left). This was followed by a 15-minute training phase in which participants experimented with the visual biofeedback to learn how to influence heart rate variability (HRV). Gloria suggests the following techniques: conscious breathing, conscious muscle relaxation, the imagination of a positively associated place, and triggering of a pleasant feeling. After this phase, the display of the biofeedback monitor (Figure 7.2) was removed and Gloria gave verbal feedback in all following exercises as soon as the HRV decreased compared to the previous value. Then, the first stress-inducing task started in which affective pictures were presented.

After a short relaxation, the second stress-inducing task started, which consisted of the Stroop Color-Word Interference Test. After this task, Gloria thanked the participants for their cooperation, said goodbye and handed them back to the experimenter.

The second training session started again with a friendly welcome and a short introduction by Gloria. The training phase was shortened to ten minutes, during which the visual feedback could be used to work on the feeling and individual technique for influencing heart rate variability. At that point, the visual feedback was again replaced by the auditory feedback from Gloria. The first stress-inducing task started in which affective pictures were presented. This was followed by a transition to the next task, during which the subjects were once again asked to consciously relax and keep an eye on their breathing. The last stress-inducing task was a five-minute stress presentation on Brexit, the withdrawal of Great Britain from the European Union. At the end of the talk, Gloria thanked them, led to a closing word and implications for everyday life, and, after saying goodbye, handed the subjects over to the experimenter again.

7.6.3.2 Stress-inducing material during the biofeedback training

For the four stress inductions, we used three methods. The first stress-inducing method was to look at pictures from two affective picture databases: 1) 47 negative pictures from the International Affective Picture System (IAPS, Lang et al. (1997)) and 2) 54 pictures of the Geneva Affective Picture Database (GAPED, Dan-Glauser and Scherer (2011)). Each picture was presented for five seconds. Participants were instructed not to look away, even if what was shown was upsetting, and to continue to try to relax and keep their HRV high or equal. If the HRV decreased, participants received encouraging feedback from Gloria.

The second stress-inducing method was a Stroop Color-Word Interference Test (Stroop, 1935), which is used as a mental stressor to analyze the mechanisms of stress reactions (Renaud & Blondin, 1997). It was shown to have stress-related effects on heart rate and HRV (Karthikeyan et al., 2013; Satish et al., 2015). The task is to name the ink color of a color word while there is a mismatch between ink color and word. Nine tables, including one tutorial table, were presented with 24 stimuli for 30 seconds each. Due to the time component of this task, the charts were displayed three seconds longer for each feedback Gloria gave.

The third stress-inducing method was to give a five-minute free presentation, in which the social evaluation pressure of the virtual observer functions as a stressor (Schneeberger et al., 2019b). The subject Brexit was chosen due to its media presence at the time of conducting the study. Participants had five minutes preparation time. After three minutes, Gloria let them know how much time was left. Gloria informed them beforehand that she would not give any feedback on the content, but would continue to give feedback on HRV.

7.6.3.3 Procedure of the diary-based stress management training

The stress diary started on day one with an introduction and explanation. In this introduction, the participants were first informed about stress in general and its effects, using the same information as in the experimental group. Afterwards, the benefit of stress diaries was explained. It was emphasized that in everyday life the triggers of stress are often not conscious and what one is not aware of, one could not work on. It promotes conscious discussion, reflection and mindfulness and brings order and overview into situations that seem overtaxing. In the following journaling part, participants first described the stressor by answering the following questions: "When did I feel stressed today?", "Describe the situation.", "What caused the stress?". After that, participants continued with analyzing their stress reaction ("How have I acted?", "How has my body reacted?", "How have I felt?"). After having created awareness for the situation, participants were asked to reflect on their own reaction and constructive strategy finding for future stress situations with the following questions: "Have I been satisfied with my reaction?", "What could I have done differently?", "How can I avoid such a situation in the future?", "How would I like to react in such a situation in the future?"

On the second day, introduction and explanation were skipped. Participants immediately arrived at the journaling part where they had to reflect on their day with the help of the same questions like on day one.

7.6.3.4 Stress-inducing material for the measurement of the dependent variables

To evaluate the effectiveness of the biofeedback training compared to the stress diary, we used an adapted version of the Trier Social Stress Test (TSST, Kirschbaum et al. (1993)). The TSST is one of the most widely used psychosocial stress tests that has been proven effective in numerous studies (Kudielka et al., 2007). The test consists of two components, a simulated job application interview and a challenging mental arithmetic task. For the first task, participants were asked to imagine that they had applied for their dream job and should now present in five minutes to the experimenter as an observer why they should get this job. The participants got informed that the experimenter will not only focus on the content but also on their non-verbal communication and that the experimenter has been trained in this respect in advance. They were also informed that their presentation would be videotaped for later analysis and recorded on a tape recorder. The participants had ten minutes for preparation, during which they were allowed to take notes. However, these were not allowed to be used during the presentation. During the participant's presentation, the experimenter acted as neutrally as possible and, if the participant stumbled or finished his presentation before the five-minute time limit, indicated the remaining time in a standardized manner. For the second task, participants were given the task of counting down the number

1022 in steps of 13. They were instructed to do this as quickly as possible and still to do it correctly. If they made a mistake, they had to start again at 1022. The task was stopped after five minutes by the experimenter. After completing the test, the experimenter informed the participant that neither the video nor sound recordings are included in the analysis and revealed the aim of the test.

7.6.4 Measurements

Demographics included gender, age, highest degree, and current engagement.

The *Perceived Stress Scale* measures the degree to which situations in one's life are appraised as stressful during the last month (Cohen, 1994). The 10-item self-assessment questionnaire uses a 5-point scale from 1 (*never*) to 5 (*very often*). The German version of this test was used for this study (Klein et al., 2016). Cronbach's Alpha was .898. The Perceived Stress Scale is used for controlling possible differences before the start of the experiment between the EG and CG.

Stress was measured before and after the tasks with the short version of the State-Trait-Stress-Inventory (Marteau & Bekker, 1992) translated in German. The STAI-6 raises the acutely felt stress with six items on a 4-point scale ranging from 1 (*not at all*) to 4 (*very*). It was measured twice: before ($\text{Stress}_{\text{pre}}$, Cronbach's Alpha .823) and after ($\text{Stress}_{\text{post}}$, Cronbach's Alpha .832) the stress-inducing with the TSST. The STAI-6 is used for hypotheses 1a, 2a, and 3a.

The *current stress* level was measured with a visual analog scale, allowing participants to make fine distinctions (Couper et al., 2006) on a continuum from 0 (*very relaxed*) to 100 (*very stressed*) in millimeter steps. It was measured twice: before ($\text{Current Stress}_{\text{pre}}$) and after ($\text{Current Stress}_{\text{post}}$) the stress-inducing with the TSST. The item was "How stressed do you feel right now?". The visual analog scale for current stress is used for hypotheses 1b, 2b, and 3a.

Task Stressfulness was measured with a visual analog scale from 0 (*not at all stressful*) to 100 (*very stressful*) in millimeter steps and should measure how stressful the stress-inducing with the TSST was perceived. The item was "How stressful did you experience the last task?". The visual analog scale for task stressfulness is used for hypotheses 2c, 3a and 3b.

Self-performance in the task was measured with a visual analog scale from 0 (*very bad*) to 100 (*very good*) in millimeter steps. The goal was to assess the self-rated performance in the TSST. The item was "How do you assess your performance in dealing with the last task?". The visual analog scale for task stressfulness is used for hypotheses 3b and 3c.

User Experience was measured with the UEQ (Laugwitz et al., 2008) on the six scales Attractiveness, Stimulation, Novelty, Efficiency, Perspicuity, and Dependability from 1 (*low*) to 5 (*high*) after both interventions.

7.7 Results

7.7.1 Subject Matter Interview

Overall, the SME rated the social biofeedback trainer agent as very valuable with extraordinary potential. Notably, the image display that allows presenting stress-inducing material was seen as a great additional benefit compared to state-of-the-art biofeedback training. The induction of stress during the training, highly increases the possibility of stress management transfer to more realistic situations. Moreover, an internalization of stress management strategies is intensified when the training is similar to everyday life situations.

Regarding the applicability of the system, the interviewed SME could imagine that the PC driven training was used not only by therapists but also in companies as a preventive program for the employees. Often, therapists had the equipment but do not have the time to advise patients during the biofeedback training sessions. Therefore, a virtual biofeedback coach could be beneficial. That a private person would buy the equipment needed for the PC driven training was assessed as rather improbable. However, our SME saw great potential for the planned mobile version of the biofeedback trainer regarding the private use-case. With smartphones and smartwatches being able to monitor the heart rate, such a mobile trainer could be “anytime and anywhere available” for users.

7.7.2 User Study

As a manipulation check, if the socially stressful task, namely the Trier Social Stress Test (TSST), was inducing stress, we compared Stress before (Stress$_{pre}$ $M = 1.90$, $SD = 0.50$) and after the TSST (Stress$_{post}$ $M = 2.26$, $SD = 0.62$). Also, we compared Current Stress before (Current Stress$_{pre}$ $M = 34.28$, $SD = 24.49$) and after (Current Stress$_{post}$ $M = 46.00$, $SD = 25.82$) fulfilling the stress-inducing task. Both, Stress ($F(1,70) = 482.89$, $p < .001$, $\eta^2_p = .87$) and Current Stress ($F(1,70) = 14.72$, $p < .001$, $\eta^2_p = .17$) significantly increased from before and to after the stress-inducing task.

To test hypotheses 1, 2 and 3c we calculated a MANOVA including all six dependent variables Stress$_{pre}$, Stress$_{post}$, Current Stress$_{pre}$, Current Stress$_{post}$, Task Stressfulness, and Self-Performance with group as factor. Overall, participants in the experimental group differed significantly from participants in the control group ($F(6,64) = 2.32$, $p < .05$, $\eta^2_p = .18$).

To show whether there is a dependency between subjective stress (Stress$_{post}$, Current Stress$_{post}$) and Task Stressfulness after the stress-inducing task and Self-performance (Hypothesis 3a and 3b) Pearson correlations were calculated.

Table 7.1: Descriptives.

	Experimental Group M (SD)	Control Group M (SD)
$\text{Stress}_{\text{pre}}$	1.79 (0.47)	2.00 (0.51)
$\text{Stress}_{\text{post}}$	2.08 (0.56)	2.43 (0.63)
Current $\text{Stress}_{\text{pre}}$	28.09 (19.99)	40.31 (27.12)
Current $\text{Stress}_{\text{post}}$	38.97 (23.51)	52.83 (26.44)
Task Stressfulness	55.51 (27.16)	66.33 (28.66)
Self-Performance	52.00 (27.26)	34.36 (24.03)
Attractiveness (UEQ)	3.61 (0.71)	3.74 (0.65)
Stimulation (UEQ)	3.44 (0.79)	3.65 (0.72)
Novelty (UEQ)	3.94 (0.86)	3.37 (0.86)
Efficiency (UEQ)	3.61 (0.60)	3.79 (0.52)
Perspicuity (UEQ)	3.79 (0.79)	3.91 (0.77)
Dependability (UEQ)	3.40 (0.62)	3.58 (0.48)

Note. $N = 71$. Stress was measured on a 4-point scale from 1 (*not at all*) to 4 (*very*), Current Stress from 0 (*very relaxed*) to 100 (*very stressed*), Task Stressfulness from 0 (*not at all stressful*) to 100 (*very stressful*), Self-Performance from 0 (*very bad*) to 100 (*very good*), UEQ from 1 (*low*) to 5 (*high*). Pre stands for the self-reported values before the Trier Social Stress Test (TSST), post for the self-reported values after the TSST.

7.7.3 Hypotheses

Hypothesis 1a stated that participants in the EG report a lower stress level, measured with the STAI-6, before the stress-inducing task than participants in the CG ($\text{Stress}_{\text{pre}}\text{EG} < \text{Stress}_{\text{pre}}\text{CG}$). We found a significant difference between EG and CG in the STAI-6 before the stress-inducing task ($F(1{,}69) = 3.24$, $p < .05$, $\eta_p^2 = .05$). Hence, hypothesis 1a was supported by our data. Thus, there was an effect of the training on participants stress.

Hypothesis 1b proposed that participants in the EG report a lower current stress level, measured with a visual analog scale, before the stress-inducing task than in the CG (Current $\text{Stress}_{\text{pre}}\text{EG} <$ Current $\text{Stress}_{\text{pre}}\text{CG}$). Our data showed that Current $\text{Stress}_{\text{pre}}$ was significantly lower in the EG ($F(1{,}69) = 4.65$, $p < .05$, $\eta_p^2 = .06$) confirming hypothesis 1b.

Hypothesis 2a stated that participants in the EG indicate a lower stress level, measured with the STAI-6, after the stress-inducing task than participants in the CG ($\text{Stress}_{\text{post}}\text{EG} < \text{Stress}_{\text{post}}\text{CG}$). We found the expected effect in our data ($F(1{,}69) = 5.92$, $p < .01$, $\eta_p^2 = .08$). Therefore, hypothesis 2a was supported by our data.

Hypothesis 2b proposed that participants in the EG indicate a lower current stress level, measured with a visual analog scale, after the stress-inducing task (Current $\text{Stress}_{\text{post}}\text{EG} <$ Current $\text{Stress}_{\text{post}}\text{CG}$). This hypothesis was supported by our data

($F(1,69) = 5.44$, $p < .05$, $\eta_p^2 = .07$).
Hypothesis 2c proposed that participants in the EG assess the stress-inducing task, measured with a visual analog scale, as less stressful than participants in the CG (Task stressfulness EG < Task stressfulness CG). The difference between the groups did not reach a significant level ($F(1,69) = 2.66$, $p = .054$, $\eta_p^2 = .04$). Thus, there was no support for hypothesis 2c.
Hypothesis 3a stated that subjective stress perception after the stress-inducing task and the assessment of one's own performance are dependent. The higher the stress level, the lower the self-assessed performance. Our data revealed that both, $\text{Stress}_{\text{post}}$ ($r = -.41$, $p < .001$) and Current $\text{Stress}_{\text{post}}$ ($r = -.39$, $p < .001$) were significantly correlated with Self-Performance. Therefore, hypothesis 3a was confirmed by our data.
Hypothesis 3b stated that task stressfulness and the assessment of one's own performance are dependent. The more stressfulness the task is experienced, the lower the self-assessed performance. Our data showed that Task stressfulness and Self-Performance were significantly correlated ($r = -.51$, $p < .001$), which confirms hypothesis 3b.
Hypothesis 3c proposed that participants in the EG assess their performance during the stress-inducing task better than participants in the control group (Self-Performance EG > Self-Performance CG). We found that Self-Performance was higher in the EG compared to the CG ($F(1,69) = 8.38$, $p < .005$, $\eta_p^2 = .11$). Hypothesis 3b was supported by our data.

Overall, these results have demonstrated that our biofeedback training seems to be more effective in reducing stress and learning stress management than stress diaries.
User Experience was analysed exploratory (Descriptives in Table 7.1). The training was rated positively regarding the usability aspects. The MANOVA including all six scales from the UEQ revealed that there is a significant difference between the groups ($F(6,64) = 3.27$, $p < .05$, $\eta_p^2 = .24$). The ANOVAS showed that, apart from Novelty, there was no significant difference between the assessment of usability between the groups. We applied a Bonferroni correction due to the exploratory approach. Novelty was assessed significantly higher for the experimental group compared to the control group ($F(1,69) = 7.96$, $p < .05$, $\eta_p^2 = .10$).

7.8 Discussion

This paper presents the first stress management training using biofeedback guided by an interactive social agent. To evaluate our approach, we conducted an interview with an expert and a user study in which we compared the novel approach against a stress management training using stress diaries. The expert interview revealed a high potential of the system as it does not require a human coach's presence for the entire time of the training. In our study, participants had to fulfill a socially stressful task after the training. We measured stress and current stress after the training, as well as stress, current stress, task stressfulness, and self-performance after a socially stressful task. Our results show that the stress levels and the self-performance rating were dependent on the training type (biofeedback vs. stress diaries). The group with the novel biofeedback training reported lower stress than the group with the stress diaries.

Several studies showed an increase in subjective relaxation in combination with HRV biofeedback training (Goessl et al., 2017; Ratanasiripong et al., 2015; Sutarto et al., 2010). In our study, we could find similar results. Participants reported lower self-assessed stress levels immediately after the biofeedback training. Thus, the HRV biofeedback training guided by the social agent seems to reduce the experienced stress level and increase the adaptability to cope with stressful situations.

Both immediately after the biofeedback training and after the socially stressful task, participants' self-assessed stress levels were lower for the biofeedback training. This might be due to the procedure of the training. In the training sessions, participants were not only exercising in a protected environment but also under increasingly stressful conditions. Therefore, it seems that participants transferred the learned strategies to other stressful situations.

Likewise before (Blouin et al., 2014), we have found that the stress level and the assessment of one's own performance are correlated. Our data shows that the higher the stress level is assessed by the participants, the lower the self-rated performance in the stress-inducing task is. Moreover, the participants in the biofeedback training group assessed their performance better than participants using stress diaries. Reduced performance might be due to the concentration difficulties, reduced memory, and learning ability when experiencing high stress (Staal, 2004).

7.8.1 Limitations and Future Work

Our biofeedback training consisted of two training sessions. Though there are studies about short duration heart rate variability biofeedback (Prinsloo et al., 2013; Prinsloo et al., 2011), maximal control of HRV can be obtained in most people after approximately four sessions of training (Sutarto et al., 2010). We could show that already, after two training sessions, participants felt significantly more relaxed. This highlights the efficiency of the training and confirms the assumption of having developed a budget training. However, the effects of more training sessions should be examined. Also, the control group, which had a diary-based stress management training, had two sessions. Journaling is a method that might be more useful for a mentally healthy population, as it has to be carried out independently. We considered this with the exclusion of participants that have a current mental illness. However, it might be that the learning process through journaling is slower and needs more practice. Moreover, the chosen study design did not allow to examine the effectiveness of the agent compared to a human coach. Therefore, we plan to design another experimental group that gets biofeedback with a human coach and compare it with the virtual coach. Doing so gives us the possibility to examine if a virtual biofeedback trainer is as helpful as a human one. Moreover, a comparison to a waiting control group without any intervention might be useful.

Regarding the missed significance for task stressfulness, a post-experimental interview with the participants revealed that several of them only evaluated the mental arithmetic task and not both the simulated job application interview and the mental arithmetic task. As participants are often more stressed by one type of the two tasks (Kirschbaum et al., 1993), the data might be not representing those participants that were more stressed by the simulated job application interview.

We measured the effect of our novel biofeedback training immediately after the second training session. Therefore, we can only assess short term relaxation effects of our social agent guided biofeedback training. To draw conclusions about long-term success, future studies should include a follow-up survey after, e.g., one month. However, it has been shown that the positive effects of HRV biofeedback training seem to be persistent after 28 days (Lemaire et al., 2011).

Based on the expert interview results, our future goal is to develop a mobile biofeedback trainer. Therefore, the components of the real-time system have to be adapted to run on a smartphone. In such a mobile setting, the sensors for measuring the physiological parameters are ideally coming with a wearable, such as a watch with a heart rate monitor.

7.9 Conclusion

The used interactive social agent technology allows sophisticated modeling of interactive social agents as mental health coaches for biofeedback training. The presented results show, once more, that such technology can be employed to assist human experts in the health context.

We developed a technology-driven stress management training with a social agent as a coach. The field of application of such a system ranges from clinical to healthy user groups. As an adjunct, it could complement state-of-the-art therapies for all mental diseases where emotion regulation and stress management are crucial, like depression, post traumatic stress disorder, or attention deficit hyperactivity disorder. Moreover, it can be used in healthy populations as a tool for preventing private and professional stress to avoid physical and psychological stress-related illnesses.

7.10 Acknowledgment

This work is funded by the German Ministry of Education and Research (BMBF) within the EmmA project (funding code 16SV8029) and the German Research Foundation (DFG) within the DEEP project (funding code 392401413). Special thanks go to our biofeedback expert Berit Greulich for taking the time supporting us during the development and evaluation of our system.

8

MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation

Patrick Gebhard[1] Tanja Schneeberger[1], Tobias Baur [2], and Elisabeth André[2]

[1]German Research Center for Artificial Intelligence
[2]Augsburg University

Published in Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018)
https://doi.org/10.1145/3397481.3450683

This chaper contains the original version (with slight changes due to formatting) of our full paper presented at the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018). During the conference, the first author presented the paper orally to the conference attendees.

Abstract

Understanding emotions of others is related to a theory of mind approach. It requires knowledge of internal appraisal and regulation processes of emotions. Multi-modal social signal classification is insufficient for understanding emotional expressions. Mainly, because many communicative emotional expressions are not directly related to internal emotional states. Moreover, the recognition of the emotional expression's direction is not considered so far. Even if social signals reveal emotional aspects, the recognition with signal classifiers cannot explain internal appraisal or regulation processes. The information the latter two provide is one approach for building cognitive empathic agents with the ability to address observations and motives in an empathic dialogue. In this paper, we introduce an emotional computational model for empathic agents. It combines a simulation of appraisal and regulation processes with a social signal interpretation that takes directions of expressions into account. Our evaluation shows that sequences of social signals can be related to emotion regulation processes. This together with appraisal and regulation knowledge enables our agent to react empathically.

Keywords: Modelling of User Emotions, Nonverbal Behavior Understanding, Empathic Agents

8.1 Motivation

Our world is a social place. Relations with others and interaction with others are essential. In many situations, we try to understand each other yet carefully managing our mental balance. Thereby, emotions seem to play a central role (Damasio, 2017). Interactive agents, such as anthropomorphic robots or virtual characters, are used for training, coaching, and assistance to help people to understand each other and develop various skills (Anderson et al., 2013; DeVault et al., 2014; Kapoor & Picard, 2005; Lisetti & Nasoz, 2002; Valstar et al., 2016a). The more agents are employed for social tasks; the more significant is the need for understanding user emotions, motivations, and related social behavior. All this can be exploited by interactive agents to adapt empathically to the user and the user's situation in general.

The crux of understanding emotions is that most, if not all, emotions are regulated internally (Gross, 2013; Tomkins, 1984). This is especially the case for emotions, such as shame, that are related to the appraisal of oneself (Lewis, 2008; Scheff & Retzinger, 2000). Only a few of the current approaches of emotion models for empathic agents take emotion regulation into account. Some of them are able to model re-appraisal processes (Dias et al., 2014; Marsella & Gratch, 2009). However, none of them explicitly combines a social signal interpretation with a cognitive modeling of appraisal and regulation processes.

Moreover, none of the existing recognition approaches considers the direction of emotional expressions. This means it is unclear to whom or what that emotional information applies. It is known from research in the area of emotional mimicry that the direction of emotional expressions is a crucial information to understand another's intention (Bourgeois & Hess, 2008; Hess & Fischer, 2013). In dyadic interactions, emotional expressions can be directed to the interaction partner, the situation, the dialog topic or at the person(s) mentioned in the utterance. By linking the gaze or head movement while observing an emotional expression, its direction can be tracked (Bänninger-Huber & Steiner, 1992; Benecke, 2002; Schwab, 2000). For example, a speaker's anger expression, directed away from the listener, provides the information that the anger is most likely addressed to something or somebody else. The knowledge about an expression's direction can be used for an automatic deduction of possible elicitors (causes) by employing different knowledge and context models. In general, the recognition of the expression's direction might be as important as the emotional expression itself, especially, if empathic agents have to generate (re-)actions based on this information.

MARSSI combines an extended social signal interpretation with a simulation of both, the appraisal and the regulation processes. The overall aim of this work is to lay the basis for a deeper analysis of social, emotional signals and their connection to cognitive processes. This may foster the widespread use of empathic agents for various assistive tasks in everyday human environments. We show a first example exploitation of our model in a job interview debriefing session. For the debriefing, a virtual character in the role of a coach addressed the observed non-verbal behavior and inferred possible appraisal and regulation hypothesis in an empathic manner.

8.2 Related Work

8.2.1 Empathic Agents

Interactive systems are more likely to be accepted if the machine is aware of the user as a social actor (Picard, 1997, p. 247). Furthermore, understanding how emotions work is key to social training applications (Johnson et al., 2000). In order to achieve this goal, recent developments in the area of empathic agents have initiated a shift from simple task-based human-machine interaction to a more human-like social interaction. Several approaches are addressing these requirements. Lester et al. (1997) and Van Mulken et al. (1998) are using virtual characters that are sensitive to the learners' emotional state to enhance their engagement and motivation. This is described as the persona effect. Bickmore (2003, p. 131 ff.) describes the interactive fitness agent Laura that was designed to build up a relationship with a human user. In order to build a working alliance, Laura uses relational strategies like giving warm facial expression. Other approaches go further and employ cognitive models of appraisal within their systems following Wilks' argument that Digital Companions must have an understanding of the human partners' emotions as a basis for a Human-Companion relationship (Bee et al., 2010, p. 4).

Conati and Maclaren (2009) present an interactive agent system that is able to model user emotions in a specific computer game. The system simulates possible user appraisals, goals, as well as motivations and models interdependencies with Bayesian networks. The emotion model uses the user's game actions as input. Rodrigues et al. (2009) propose a generic computational model of empathy. In their model, they implement a reactive perception of others' affective state and the subsequent generation of an empathic response. However, the authors focus on the empathy between virtual agents and not between an agent and a user. Dias et al. (2014) present FAtiMA, a generic and flexible architecture for emotional agents. It supports re-appraisal processes and the use of theory of mind models. How re-appraisal processes are interfering with internal situational representation is not explained.

One of the most powerful computational models of emotions is EMA. It is used by empathic agents in various systems (Swartout et al., 2006), for example, to model appraisal and reappraisal of users (Marsella & Gratch, 2014). Like in the previously mentioned work, goals and motivations are represented. In addition to that, EMA provides an explicit representation of coping strategies that can also be used to model a user's situational coping. Albeit coping mechanisms are related to the emotion regulation process, they differ conceptually. As a result, EMA does not allow explicit modeling of complex social emotions like shame. Also, it is unclear how to relate observed social signals to re-appraisal processes.

Looking at state-of-the-art computational models of user emotions for agents, it becomes clear that essential concepts like emotion regulation, emotional expressions direction, as well as relations to sequences of social signals, are neglected.

8.2.2 Emotion Modelling and Theory of Mind

Computer scientists focus on cognitive appraisal theories for emotions (Moors et al., 2013). Because of their concept of modeling processes and signals they can be realized in computer programs. The computational modeling of emotions started in the 1980s (Pfeifer, 1988) and is continuously refined (Marsella et al., 2010; Rodrıéguez & Ramos, 2014). Psychological theories of appraisal rely on a particular input, such as, goal information, certainty, situational control, and the elicitor (who or what is the cause). Additionally, the appraisal might rely on information from a theory of mind (ToM) of others that represents hypotheses about another's mental states, status, and role (Leudar et al., 2004; Premack & Woodruff, 1978). The outcome of the appraisal process is situational information, labeled with emotion term(s). According to the mentioned theories, elicited emotions influence behavior described with action tendencies (Frijda, 1987), scripts (Tomkins, 1984), or facial or vocal expressions (C. A. Smith & Ellsworth, 1985). Alternatively, more general, emotions are linked to behavioral patterns how to cope with the situation (Lazarus, 1991).

Computational models realizing such theories are used to create believable behavior of virtual characters (Vinayagamoorthy et al., 2006). Besides, they can be used to model user's appraisal(s) in a situation. A verification of the modeled appraisal information (e.g., unexpectedness) can be realized with signal-based emotion recognition (e.g., raised eyebrow), as suggested by the psychologists Mortillaro et al. (2012). However, none of the current computational models of emotion provides this.

Currently, automatic model-based emotion recognition focuses emotional expressions and related features in voice, face, gestures, and body movements (Sec. 8.2.3). The essential information to whom or to what the emotion is directed, the *emotion target*, is not included in current recognition processes. Knowing, for example, that a communicated negative emotion (e.g., anger or disgust) is not directed to an interaction partner might be a relief for that partner. The results of a study by Merten (1996) suggest that the aversion of gaze (by the sender) while communicating a negative emotion lets the interaction partner know this information is not directed to her/him. Also, current approaches do not consider the function of communicative emotions "[...] in dyadic interactions, as there are the speech-illustrating function [cf. (Bavelas & Chovil, 1997)], the function of emotional expression, and relationship-regulation" (Merten, 2003). Our model-based approach of recognizing emotional expressions takes the user's gaze and head movements into account in order to derive the emotion's target and to relate possible elicitors. Moreover, we show that the target information is central to the analysis of social signals related to emotion regulation processes.

There are few ideas in the computational realization of emotion regulation processes, mainly based on the motivation that they are an existential part of a human's emotion management. Some of the current ToM-based computational models of emotions can represent basic regulation rules (as re-appraisal rules) but not complex social emotions, such as embarrassment (Marsella & Gratch, 2014). Also, none of the existing computational models of emotions include a real-time social signal-based emotion regulation recognition.

Recently, there are interdisciplinary approaches for computational models of emotions aiming to bridge the gap between modeled emotions and actual user emotions. One of the latest attempts employs a ToM of user emotional states in a social job interview simulation (Belkaid & Sabouret, 2014; Youssef et al., 2014). Using belief, desire, and intension (BDI) rules (Rao & Georgeff, 1995), three categories of user mental states are modeled: intentions, beliefs, and emotions. The quality of social relations is based on liking and dominance values. The input of the model is the illocutionary part of speech acts (speaker intention).

The model is embedded in a job interview simulation and helps to improve the system's training efficiency. A corroboration of modeled appraisal information with a real-time social signal analysis is not included.

To conclude, most of the current computational models of emotions follow the concept of cognitive appraisal-based emotion elicitation. With all existing approaches, the primary challenge remains: building a probabilistic model that relates observed social signals to possible situational appraisal regulation representations.

8.2.3 Social Signal Interpretation

Social signal analysis is known to be a very hard problem and a real bottleneck in social human-agent interaction. Traditionally, research has concentrated on posteriori analyses of prototypical social cues under laboratory-like conditions. Such an approach leads, however, to over-optimistic assessments of recognition rates that cannot be re-produced in naturalistic settings. A typical example includes voice data from actors for which developers of emotion recognition systems reported surprisingly high accuracy rates of nearly 80% for a seven-class problem. When moving to more naturalistic scenarios, such as child-robot interaction, accuracy rates went down considerably to about 40% for a five-class problem. An experiment that compared relevant features and recognition rates for acted and spontaneous emotions has been conducted. The experiment revealed that adequate segment lengths and relevant features could not be transferred from acted to spontaneous emotions (Vogt & André, 2005).

An obvious approach to improve the robustness of the analysis is the integration of data from multiple channels. A meta-study on 30 published studies of multimodal affect detection comes to the interesting conclusion that performance improvement, i.e., the improvement of the fused decisions compared to the best unimodal classification, correlates significantly with the naturalness of the underlying corpus (D'Mello & Kory, 2012). While an overall mean multimodal effect of 8.12% is reported, they also found that improvements are three times lower when classifiers are trained on natural or semi-natural data (4.39%) compared to acted data (12.1%). At first glance, the meta-study suggests that under realistic conditions there is less room for improvements than in the case of acted material. However, when analyzing the investigated approaches in more detail, it becomes apparent that most of these approaches make unrealistic assumptions, which are hard to meet in real-life environments. Therefore, they do not achieve the expected improvements as different channels are combined with fixed time segments, e.g., between the beginning and the end of an utterance. It has the drawback that cues from other modalities outside the segment will be missed. Promising approaches to overcome these limitations include the use of Multi-stream Fused Hidden Markov Models (Zeng et al., 2008) as well as Multidimensional Dynamic Time Warping

(Wöllmer et al., 2009).

Furthermore, attempts have been made to improve recognition rates by taking into account the dynamics of social signals. A person showing signs of happiness (usually) will not fall into a deep depression within the next few seconds. Taking the temporal context into account allows building models that are less prone to false detections. Fusion architectures based on Hidden Markov Models and Dynamic Bayesian Networks appear to be very suitable to model how social signals evolve over time. More sophisticated approaches, such as bidirectional Long Short-Term Memory (Wöllmer et al., 2013), add more flexibility to the fusion process by learning the optimum amount of context to be taken into account.

The fusion processes mentioned above consider the temporal history of social signals. However, they do not consider the context of the social signals. So far, emotions are analyzed in isolation without considering the emotion-eliciting stimuli. This is extremely hard if not impossible (Kächele et al., 2015; Parkinson & Manstead, 2015). For example, a smile is not always a sign of happiness. People also tend to smile when feeling embarrassment (Keltner, 1995). Furthermore, how emotions are perceived depends on the social relationship between interlocutors (Ursula & Shlomo, 2015), for example, a person may interpret a smile of a competitor rather as gloating. Many recognition systems are not able to take these subtle differences into account. Rather they would map a smile onto the emotional state happiness. First attempts to the situational context for emotions are made by using a probabilistic framework (Conati & Maclaren, 2009). However, this work focuses on the prediction of emotions from the situated context while the potential of external signs of emotions has not been fully exploited.

A recent study conducted by de Melo et al. (2014) analyzed the behavior of people engaged in the prisoner's dilemma with counterparts and found out that people derive information from appraisal processes when analyzing the emotional displays of others. Their study reveals the importance of appraisal-based models for the interpretation of social and emotional cues. This insight is shared by Mortillaro et al. (2012). Based on the observation that current emotion recognition systems use a so-called 'black-box' approach that map low-level features onto abstract emotion labels following statistical methods, they advocate the use of appraisal-based models to guide emotion recognition tasks. In particular, they propose appraisals as an intermediate layer between social cues and emotion labels. Nevertheless, neither the direction of emotional expressions are included, nor does the model include an estimation of emotion regulation strategies based on social cues.

8.3 Required Concepts

Clark and Krych (2004) point out that the observation of human social signals is mandatory for a mutual understanding of a dialog partner. In line with this view is the computational model of emotional grounding (Bosma & André, 2004) that helps to identify the user's intention in a natural language dialogue by relying on the users's emotional state. In comparison to Conati (2002), we consider not only the emotional signals by the users but also the cause of emotions. However, both approaches did not clearly distinguish the emotion origin, such as an internal, related to a person's self, emotion (*structural emotion*), a result of the appraisal of a situation (*situational emotion*), or an emotional message expressed non-verbally (*communicative emotion*) (Moser, 2009, p. 111-112). This classification schema has not found its way into computational models of emotions and approaches for recognizing emotions yet.

The combination of a social signal interpretation with modeling of structural, communicative, and situational emotions can be used to build a differentiated, probabilistic model of user's emotional states during dialogue. This approach requires a representation of (mostly) unconscious relevant processes and mental states that build a foundation for an empathic dialogue with users.

A unique, rarely by a computational model of affect included, aspect concerning structural emotions is the - mostly unconscious - regulation of *intrapersonal* emotions (Gross, 2013, p. 6; Tamir, 2011). In that process, cultural and individual emotion regulation rules might inhibit or alter elicited structural emotions. A cognitive emotion appraisal concept extended by regulation rules enables a simulation of various adapted, or inhibited emotions. Notably, the regulation process can be related to social signals (Bänninger-Huber et al., 1990; Benecke, 2002; Moser & von Zeppelin, 1996; Schwab, 2000), which can be recognized by a real-time social signal interpretation component. No current computational approach of emotion recognition take regulation processes and related social signals into account. Both, their importance and necessity for understanding human emotions are described by cognitive psychoanalysts (Moser & von Zeppelin, 1991, 1996). Relying on the combination of regulation processes and social signals for emotion recognition is of particular importance when considering that the mapping of emotional expression (even considering the fusion of several modalities) onto emotional states is not reliable (Hess & Fischer, 2013; Kächele et al., 2015; Kaiser & Wehrle, 2001; Parkinson & Manstead, 2015; von Scheve, 2010).

8.3.1 Structural, Situational, and Communicative Emotions

In 1990, psychologists introduced a ToM concept of how to combine an offline social signal interpretation with modeled emotions and emotion regulation processes (Bänninger-Huber et al., 1990). The work aimed at the creation of an emotion regulation process model. Based on this, Moser and von Zeppelin (1996) designed a theory of emotions that differentiates between *communicative emotions*, *structural emotions*, and *situational emotions*.

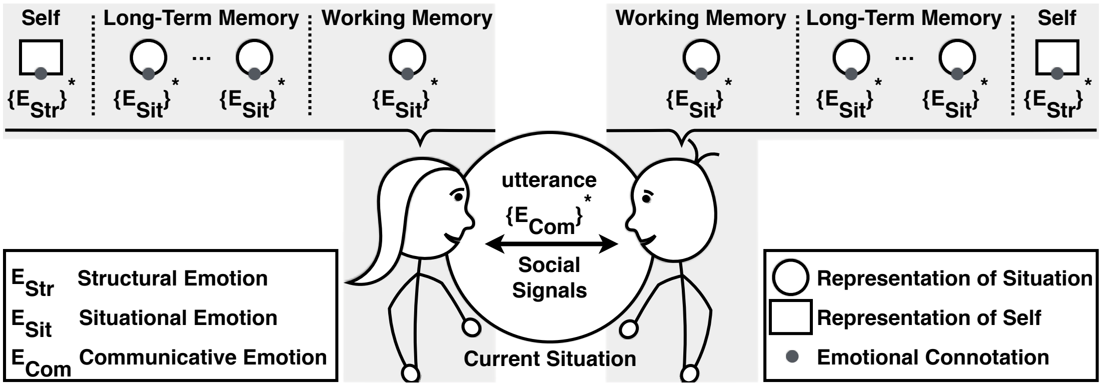


Figure 8.1: Structural emotions, situational emotions, and communicative emotions in a dyadic interaction setup.

This *functional classification of emotions* helps to describe emotions and their implications on internal processes as well as their reflection in behavior more distinguishable:

Structural emotions represent information about the appraisal of oneself and hence are related to the self-image (Figure 8.1, top, left and right). Such emotions are, for example, shame, pride or gratitude.

Situational emotions represent information that is linked to a topic or situation that have been experienced (Figure 8.1, top, center, long-term and working memory). Situational emotions reflect the level of security. More specific, such emotions like fear or distress reflect the fact that the situation comes with unforeseen or unbearable requirements. If a situation addresses social skills or relations, the emotions shame or pride might be linked.

Communicative emotions are encoded non-verbally in *sequences of social signals*, like in vocal or facial expressions (Figure 8.1, center). They are, for example, described by Ekman (1992). "Communicative affects bring the regulatory systems [and related structural, and situational emotions, author's remark] of both interaction partners in relation and they provide rapid information about the partner's regulatory state." (Moser & von Zeppelin, 1996, p. 111). One of the most crucial aspects of communicative emotions is that they are directed towards the dialog partner or situational objects (Bänninger-Huber, 1996; Schwab, 2000). The class of communicative emotions includes social signals that are used for relationship regulation/management (esp. smile, Bänninger-Huber, 1996, p. 72 ff.), which is related to social mimicry processes (Hess & Fischer, 2013; Lakin et al., 2003).

8.3.2 Emotion Regulation

An emerging research focus on cognitive emotion theories is the *regulation of emotions* (Gross, 2013). Tomkins proposed that adult emotions are almost always regulated (Tomkins, 1984). The regulation of emotions describes the process of suppressing or changing emotions if they do not fit the current individual situation. The main purpose of the regulation process is to "cover" an unwanted emotion with others in order to (re-)establish the feeling of being secure (Tamir, 2011).

The regulation process changes the situational appraisal information, which elicits a different emotion reflecting a "better" (with regard to the individual's situational appraisal) management (coping) of the situation. The employed *regulation strategy* changes situational values of individuals' internal situational representation in the working memory (Figure 8.1, top). Classes of situational changes are described by Moser (2009, p. 39): 1) *actor transformations (self as actor → other as actor, other as actor → self as actor)*, 2) *action transformations (e.g., action → opposite of action, action → denial of action)*, and 3) *object transformations (object x → self as object, object x → y as object, x≠y, y≠self)*. As a result, an individual situational representation differs from the current outside situation. This view explains different individual situational descriptions. With our approach, we follow the suggestion that the regulation of emotion should be part of any appraisal process model (Moors et al., 2013).

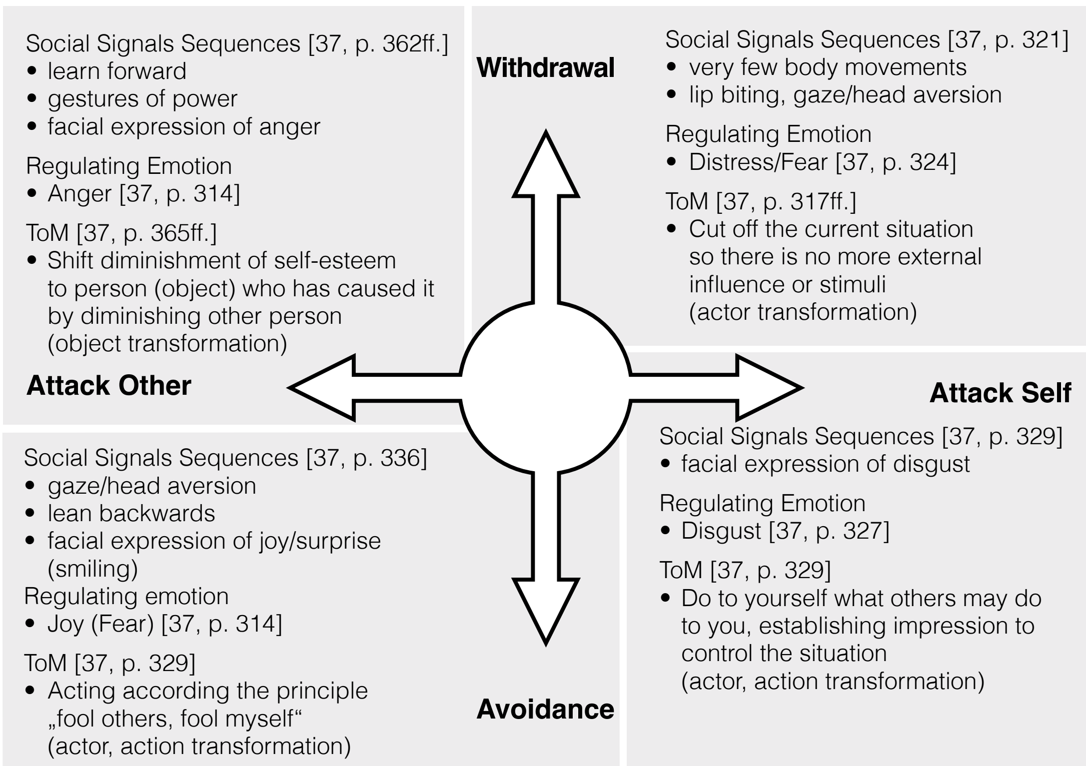


Figure 8.2: Possible shame regulation strategies, related sequences of social signals, and explanation examples.

There is evidence that the regulation process can be observed through related social signals (Bänninger-Huber, 1996; Benecke, 2002; Moser & von Zeppelin, 1996; Nathanson, 1994; Schwab, 2000). For building a computational ToM model for the structural emotion shame, we rely on a shame regulation model (Nathanson, 1994). It is a model that takes 1) clinical observations, 2) individual background and information about personal motivations, and 3) typical sequences of social signals of emotion regulation into account. For the regulation of the structural emotion shame, Nathanson describes four strategies with related social signals, and regulating emotions: 1) *Avoidance*, 2) *Attack Self*, 3) *Attack Other*, and 4) *Withdrawal* (Figure 8.2). Most likely, regulating emotions are expressed (as a communicative emotion) in the sequence of social signals that is related to individually chosen regulation strategy.

For example, *Withdrawal* is accompanied by head adaptors, lip biting, slight body movements, or avert head/gaze, *Avoidance* is accompanied by averting head/gaze or gaze wandering. Social signals are indicating a regulation process sometimes differ only minimally. For *Attack Other*, related social signals are directed gaze, spacious gestures/posture. Both, 1) the social signals of the regulation process (while processing the regulation strategy), and 2) the social signals of the regulating emotion compose identifiable signal patterns. These patterns allow conclusions to be drawn on the regulatory process and strategy. In the case of *Avoidance*, the regulating emotion is joy (triggered by the concept *"fool others fool myself"*, Nathanson, 1994, p. 339) with the corresponding facial expression smile. These signal sequences can be detected and interpreted in real-time by the MARSSI's social signal interpretation component. A result is an increased accuracy for recognizing structural emotions.

8.4 MARSSI

This section discusses required knowledge representation, the components, and the overall workflow of MARSSI. The simulation of possible user emotions relies on cognitive modeling of appraisal rules, emotion regulation rules, and social signal classifiers. The latter requires real-time signal data from an eye tracker for capturing eye movement, a depth camera for capturing head movement, facial expression, gestures, and posture; and a microphone for voice.

8.4.1 Emotion Classes, Rules, and Classifiers

MARSSI extends the emotion types from Ortony, Clore, and Collins (OCC) by Moser's and von Zeppelin's functional emotion classification (Sec. 8.3.1). All OCC emotions are assigned to the functional emotion class situational emotion, except the emotions of the types *Attribution* and *Well-Being/Attribution*. They are assigned to the functional class structural emotions since they are related to the self-image.

An *Appraisal Rule* defines how a situation is judged. With regard to cognitive appraisal theories, the situation is the elicitor of emotion. An appraisal rule represents how a user would appraise a situation. Multiple appraisals are allowed. We rely on the OCC appraisal theory (Ortony et al., 1988) with its implementation by A Layered Model of Affect (ALMA; Gebhard et al., 2003; Gebhard, 2005), e.g., *GoodActSelf*→ {*agency=self, praiseworthiness=1.0*}. In this work, we use ALMA's appraisal tag representation, like GoodActSelf, to describe an appraisal. In this case, the tag is a shortcut to the reasoning process in which appraisal rules infer a positive praiseworthiness of the action regarding the agent's goals, current situation, and related facts. MARSSI extends the appraisal notation with a confidence value representing a value how likely the appraisal fits the detected social signals. The value is computed by social signal classifiers.

A *Regulation Rule* defines how an internal emotion is regulated by changing the current appraisal information triggering a re-appraisal process that elicits a regulating emotion. Regulation rules are used to model how a user might regulate internal emotions. Multiple regulations are allowed. MARSSI extends ALMA by processing regulation rules (Sec. 8.3.2). We created regulation rules for the structural emotion shame following Nathanson's regulation theory (Figure 8.2). All regulation rules contain *situational change rules* (marked with *sit_chg*) and corresponding OCC appraisal information: 1) *AttackOther*→ {*sit_chg:object self*→ *object other; agency = other, praiseworthiness = -1.0*}. This rule regulates shame with reproach, elicited by a negative praiseworthiness by shifting the appraisal focus from one own's flaw to a blameworthy action of the person who is responsible for the shame experience. 2) *Withdrawal*→ {*sit_chg:other as actor*→ *self as actor; agency = self, desirability = -1.0*}. This rule regulates shame with distress, elicited by a negative desirability but replacing the person who is responsible for the shame experience with oneself, to the purpose of having control over the situation. A similar Withdrawal rule might include a negative likelihood to elicit the regulating emotion fear. 3) *Avoidance*→ {*sit_chg:action*→ *opposite of action*|*denial of action*|*...; agency = self, desirability = 1.0*}. This rule regulates shame with joy, elicited by a positive desirability of the imagined positive event in which the shame action has not happened. 4) *AttackSelf*→ {*sit_chg:other as actor*→ *self as actor, action*→ *intellectualization of action; agency = self, liking = -1.0*}. This rule regulates shame with disgust, elicited by a negative liking and the transformation of the shameful action into an own "ugly" character feature that is less intense and can be changed by oneself in the future. Because the person who is responsible for the shame experience is replaced with oneself implicates having control over the situation. All regulating emotions of the shame regulation rules are situational emotions that are most likely communicated (non-)verbally (e.g., Nathanson, 1994, p. 315 ff.), hence become communicative emotions. Note that each regulation rule's OCC variable hold the maximal value (e.g., 1.0 or -1.0). Its sign determines the type of emotion. Its value can be used to calculate an emotion's intensity. Currently, we are interested in the type only. Each rule holds a confidence value that is computed by social signal classifiers during runtime, representing a value how likely the regulation fits the detected social signals.

Social Signal Classifiers in MARSSI are conceptually related to appraisal and regulation information expressed as communicative emotions. We employ classifiers that are able to detect *sequences of social signals* as they occur in the situation of emotion regulation. We focus on classifiers for head (gaze), specific gestures, and posture changes for the following appraisal and regulation information: 1) *BadEvent*: user expresses anger directed towards the situation - away from the dialog partner, 2) *BadActOther*: user expresses anger towards the dialog partner, 3) *BadActSelf*: user shows facial expression of shame (e.g., blushing), head/gaze points downwards, posture is slumped down, for all shame regulation classifiers: the regulation takes time and might be accompanied by 4) *BadActSelf*→ *AttackOther*: a lean forward posture and gestures that take up room, and the user expresses anger towards the dialog partner, 5) *BadActSelf*→ *Avoidance*: a lean back posture, gaze and head aversion, and the user expresses joy towards the dialog partner, 6) *BadActSelf*→ *Withdrawal*: few body movements, gaze/aversion, and the user expresses fear away from the dialog partner, 7) *BadActSelf*→ *AttackSelf*: expresses disgust away from the dialog partner, head/gaze is mainly pointed downward.

To this end, the models for recognizing single social cues included in MARSSI are trained using machine-learning supported annotation tool NOVA[1]. To fuse multiple social signals, we employ *Dynamic Bayesian Networks* (Murphy, 2002). One of their main advantages is that they allow theory-based modeling of the structure and relevant features (represented by nodes) of a higher-level concept (e.g., regulation of shame with withdrawal), but the probability distribution of single nodes may be learned from data. Further Dynamic Bayesian Networks (DBNs) support the concept of time, allowing to model and learn temporal sequences for the interpretation of social signals. We first employ multiple classifiers trained to predict single social cues (such as facial expressions, gaze direction) to create automated annotations. For each situation, human experts manually label higher-level concepts, such as the emotion regulation strategies (Sec. 8.5).

During run-time, a confidence value, computed by the output of the nonverbal interpretation of the appraisal and regulation strategy is forwarded to the emotion simulation component, updating the possibilities of each modeled appraisal and regulation information.

[1] github.com/hcmlab/nova

8.4.2 Components and Workflow

Figure 8.3 shows how MARSSI (bottom) extends a typical appraisal approach (top) illustrating the components and workflow. Both approaches are extended by a Social Signal Interpretation component.

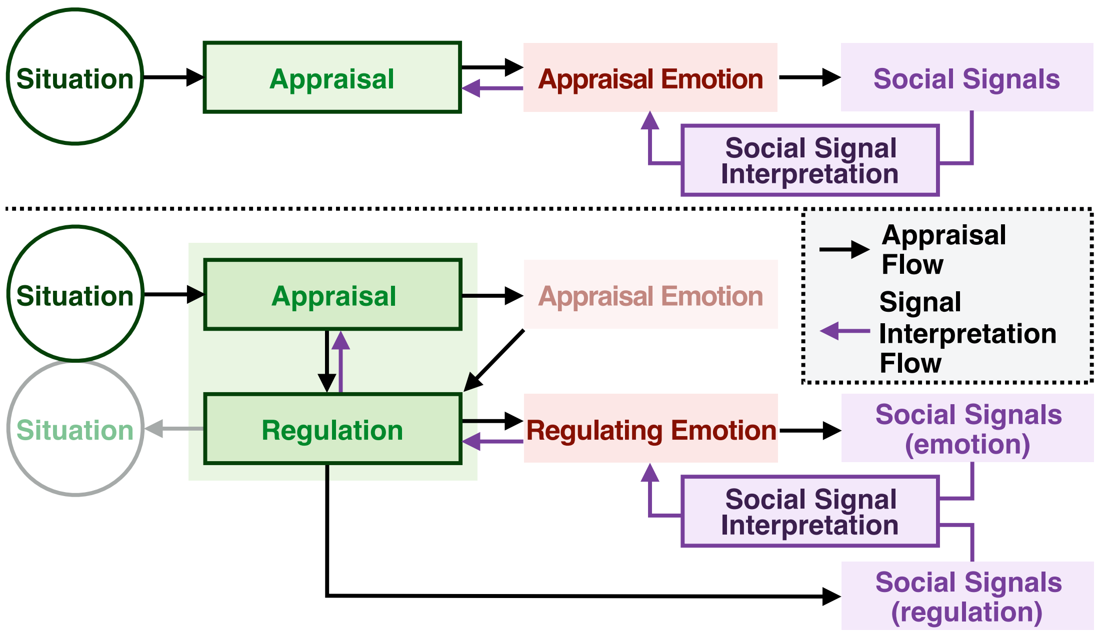


Figure 8.3: Typical cognitive appraisal process flow (top), MARSSI appraisal and regulation flow (bottom).

The MARSSI user emotion simulation is based on ALMA (Gebhard, 2005) and the Social Signal Interpretation framework (SSI; Wagner et al., 2013). ALMA provides a flexible appraisal interface and is able to simulate multiple emotional states in parallel. It was extended straightforwardly by the required regulation process and required confidence representations for appraisal and regulation representation. SSI especially allows the synchronized processing of multiple sensor inputs in real-time. This includes the extraction of relevant features at runtime and the appliance of machine learning models, such as deep neural networks or support vector machines (SVM) for predicting single cues, such as changes in gaze direction, facial expressions, gestures, and postures.

Our simulation of user emotions is structured according to conceptually coherent situations in dyadic interactions (e.g., question-answer, or comment) between a speaker and a listener. Technically, we rely on a voice signal analysis (plus gaze and head movement detection) to infer the dialog partner's attention, and actions (e.g., a user starts/stops speaking) implemented as SSI classifiers (Baur et al., 2015). The speaker is supposed to ask an emotion triggering question. While the speaker starts asking the question, the simulation of the listener's emotions is prepared (preparation phase), and the signal recognition is activated (recognition phase).

The preparation phase triggers the actual emotion simulation by a *set of appraisal and regulation annotation* given as input (e.g., {*([BadActSelf], [AttackOther, Avoidance, Withdrawal, AttackSelf])*}). Currently, the annotation is provided by human experts that annotate the situation with that specific information (Sec. 8.5). The annotation could, theoretically, derived automatically having a full-blown ToM of that specific user. In this work, we focus on the simulation of the interconnections between appraisal, regulation, and social signals (Sec. 8.3.1). Each appraisal and regulation rule input let MARSSI create a separate *emotion simulation session* (*emo_ss*). The example input creates five emo_ss, each holding appraisal information, the *elicited emotion*, and (if a regulation rule is stated) the *regulation rule*, and the *regulating emotion*: 1) (*BadActSelf→ Shame*), 2) (*BadActSelf→ Shame→ AttackOther→ Reproach*), 3) (*BadActSelf→ Shame→ Avoidance→ Distress*), 4) (*BadActSelf→ Shame→ Withdrawal→ Joy*), 5) (*BadActSelf→ Shame→ AttackSelf→ Disgust*).

The recognition phase lasts as long as the listener handles the question or the comment. Within that phase, the Social Signal Interpretation updates the appraisal and regulation confidence values in each emo_ss reflecting the match of detected social signals to the appraisal and regulation information in each emo_ss.

8.5 Evaluation and Example Simulation

This section explains how we employed MARSSI for an empathic agent. First, we need recorded data of participants in specific situations that elicit the structural emotion shame to build our corpus. We used a job interview situation and tried to elicit the structural emotion shame in the interviewees. To generate shame eliciting situations, we conducted a pre-study. Two job coaching experts identified six possible shame eliciting situations considering Nathanson's work (Sec. 8.3.2). 26 participants (age 18 - 29, M = 21.71, SD = 2.91) were asked to put themselves into a position of a job applicant experiencing these six different situations. The task of the participants was to describe in their own words how they would react. The answers were analyzed by two psychologists and assigned to Nathanson's four shame regulation strategies (Figure 8.2). Finally, we identified five situations that elicit the structural emotion shame, such as, *"Before we begin, let me ask a short question: Where did you find your outfit? It really doesn't suit you."*

To generate our corpus, we created a 15min job interview with the five shame eliciting situations from the pre-study. In our evaluation, this job interview was conducted by a female interviewer with 20 participants (10 female, age 19 - 30, M = 24.60, SD = 4.08) as a role-play. After welcoming the participants, they were asked to imagine that they applied for a student assistant job in their favorite faculty. Each participant is sent to the interviewer's office for a job interview. Afterwards, the participant answered demographic questions and was compensated. The interviews were recorded with a depth camera and a head-mounted microphone.

In total, 100 (20 participants in five situations) shame eliciting situations are building the corpus for the analysis. We annotated the obtained data in order to create the social signal classifiers. Each situation was classified independently by three students, that were not related to the experiment neither knew about the aim of the study. They were trained beforehand to classify Nathanson's four shame regulation strategies. Overall, 300 labels were assigned as follows: 83 Withdrawal, 105 Attack Self, 98 Avoidance and 14 Attack Other. For assessing the reliability of agreement Fleiss' kappa was calculated for three raters, four labels, and 100 data points. With 0.7301 it is considered as substantial agreement.

Based on this data, we trained the Bayesian network in a 50:50 split validation approach. To this end, we employed several social signal processing algorithms to generate labels for single social cues on multiple modalities of both the interviewer and the candidate. Some cues are calculated based on single, meaningful features, such as the energy of the motion vectors of both hands of a participant or the overall movement of the hands, head touches, and the openness of the body posture (Baur et al., 2015).

For more complex cues, suvh as subtle smiles, we employed an SVM to train models based on manual annotations on the training subset of our corpus. For cues related to the head and face, we thereby extracted OPENFACE (Baltrušaitis et al., 2016) features. Analogously, we repeated this step for other modalities, such as the paralinguistic channel, by training a model to detect spoken words, fillers, and silence, as well as models to detect the level of arousal from the audio modality based on GEMAPS (Vogt et al., 2008) features. A human annotator interactively corrected the annotations when necessary, and after each session, the models have been retrained as proposed in (Wagner et al., 2018b).

To find the ground truth of the observed emotion regulation strategy, we additionally labeled time segments including the duration of each question and the candidate's answer, with 1) the type of question as additional context information and 2) with the rating of human labelers for the classes related to regulation cues (e.g., AttackOther, AttackSelf, Avoidance, Withdrawal, and None).

Finally, based on these semi-automated annotations we created a training set. It contains the parallel appearance of the ground truth labels for the shame emotion regulation strategy, the context information and the single observed social cues (we discretized continuous annotations) and trained a DBN using the Expectation Maximization algorithm, to learn both the distribution of the single labels in our corpus, but also their influence on the single shame regulation strategies. Overall, the network achieved a precision of 82% for Avoidance, 65% for AttackSelf and 64% for Withdrawal from non-verbal behaviors only. The training data provides too few social signals related to the AttackOther strategy. As a result, the DBN could not be trained to that extend.

In a next step, we used the cognitive modeling and the trained social signal classifiers to simulate user emotions in real-time in a debriefing session with our interactive virtual character Tom. He has the role of a coach discussing the user's (non-verbal) reaction to the interviewer's question. Tom is embedded in a 3d virtual environment (Figure 8.5) capable of performing social cue-based interaction with the user. He is able to perform lip-sync speech output using the state-of-the-art Nuance Text-To-Speech system. Tom comes with 36 conversational motion-captured gestures and has 14 facial expressions including the six basic emotion expressions.

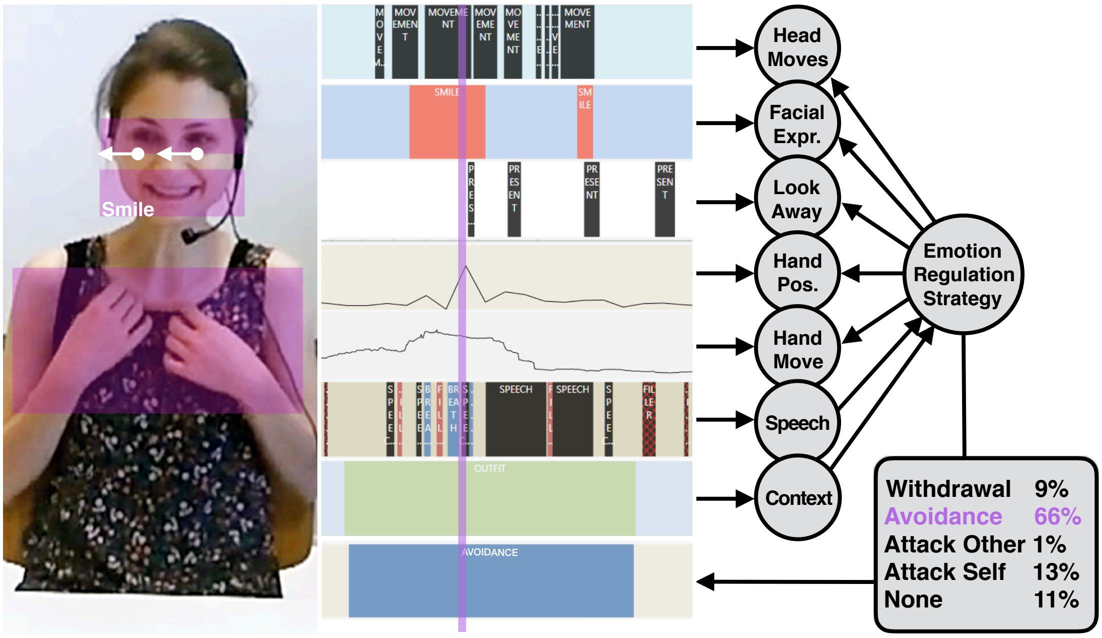
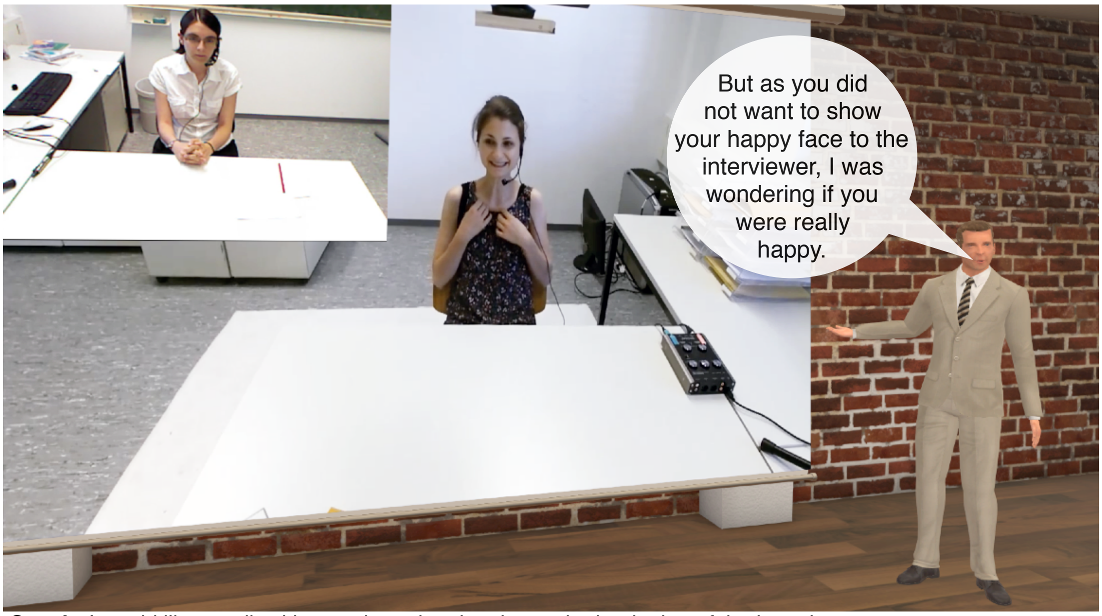


Figure 8.4: Recognized and annotated cues are fed in a DBN that infers the current shame regulation strategy and predicts it in real-time.

For each shame question, possible appraisals and regulations of the applicant were prepared by MARSSI. Each preparation phase (Sec. 8.4.2) is triggered by the voice activity signal of the job interviewer, posing the question. In fact, the following appraisal/regulation input is given to MARSSI for each shame question: {*([BadEvent]), [BadActOther], ([BadActSelf], [AttackOther, Avoidance, Withdrawal, AttackSelf])*} with *[BadEvent]* denotes the appraisal that the situation as noisy, *[BadActOther]* denotes the appraisal that the interviewer's action is blameworthy, e.g., the interviewer speaks with an inappropriate low voice, and *[BadActSelf]* denotes the appraisal that the question triggers a blameworthy memory of the applicant. The latter elicits the structural emotion shame that the applicant most likely will regulate with the 4 mentioned regulation strategies (Sec. 8.3.2 and Sec. 8.4.1). As a result seven emo_ss (Sec. 8.4.2) are created holding appraisal information, the *elicited emotion*, and (if a regulation rule is stated) the *regulation rule*, and the *regulating emotion*: 1) (*BadEvent*→ *Distress*), 2) (*BadActOther*→ *Reproach*), 3) (*BadActSelf*→ *Shame*), 4) (*BadActSelf*→ *Shame*→ *AttackOther*→ *Reproach*), 5) (*BadActSelf*→ *Shame*→ *Avoidance*→ *Distress*), 6) (*BadActSelf*→ *Shame*→ *Withdrawal*→ *Joy*), 7) (*BadActSelf*→ *Shame*→ *AttackSelf*→ *Disgust*). At the same time, the related social signal classifiers are activated (Sec. 8.4.1). At runtime, the confidence values from the classifiers update appraisal and regulation representations (Figure 8.4).



Coach: I would like to talk with you about the situation at the beginning of the interview. The interviewer commented on your outfit. Is this ok with you?

User: Sure.

Coach: Do you first want to see the video from the interviewer's position?

User: Yes.

[system plays the recorded video, pauses three times, coach explains …]

Coach: In this situation, the interviewer was attacking your outfit saying that it does not fit you. As you know, I kept a watch on your facial expression and your body language during the interview. I could observe that you were smiling and looking away from the interviewer while answering.

1

Coach: It seems like you did not want to look at the interviewer anymore though you were smiling. Because of the smile, I could have thought you were happy first. But as you did not want to show your happy face to the interviewer, I was wondering if you were really happy. Maybe the attack on your appearance made you feel bad, but you did not want to show it. That is ok.

2

Coach: To defend themselves, others sometimes do not at all understand the attack but think the interviewer said their outfit fitted nicely. If someone said my suit didn't look good, I also would feel hurt. But don't worry, the interviewer just said this to get you off your feet, because you are already at the advanced level of the training.

3

Figure 8.5: Virtual coach discusses prominent situations.

Our empathic agent exploits MARSSI's knowledge of the appraisal and the regulation strategies in order to generate an empathic reaction. Currently, the reaction is based on the detected appraisal or regulation with the highest confidence value. The aim is to support in the user's self-reflection by explaining to her what MARSSI discovered from the social signals. We elucidate this with the example of the regulation strategy Avoidance. Avoidance is one of the four regulation strategies when experiencing the structural emotion shame (Nathanson, 1994). It is accompanied by specific facial expressions and body language (Sec. 8.3.2). This strategy can also be expressed verbally by redirecting the subject to another. We focus on the facial expression and body language. In general, Tom (Figure 8.5, right) would first explain what social signals MARSSI have detected and which regulating emotions are related. Afterwards, he would subtly explain the connection to the underlying structural emotion. We want to outline a possible interaction between a user and the coach where MARSSI detected the following rule *Avoidance*→ {*sit_chg:action*→ *opposite of action*|*denial of action*|*...; agency = self, desirability = 1.0*} in the example situation with the interviewer *"Before we begin, let me ask a short question: Where did you find your outfit? It really doesn't suit you."* This rule regulates shame with joy, elicited by a desirable imagined positive event in which the shame action has not happened.

As seen in Tom's explanation, he does not directly address the structural emotion. Especially in those cases where the underlying structural emotion might be shame, the subtle approach is extremely important. Since shame is the emotion that is connected to the evaluation of the self, the coach has to be very sensitive such that the user is still able to preserve his self (Lewis, 2008; Scheff & Retzinger, 2000).

In the example situation, MARSSI recognized the regulation strategy Avoidance. We generate the explanations with textual templates for: 1) situation description (and for the first shame question, explanations of Tom's role) and found social signal sequences related to appraisal and regulation strategies (Figure 8.5, 1), 2) general explanation how such signals could have interpreted (Figure 8.5, 2), and 3) explanation of the regulation process and typical observations (Figure 8.5, 3), which we took from descriptions of Nathanson (1994, p. 303 ff.) and the two coaching experts.

8.6 Conclusion and Future Work

In this paper, we have presented the computational model of emotion MARSSI that relates appraisal rules and emotion regulation rules with social signal interpretation. MARSSI employs an extended theory of emotions that comes with three functional dimensions to emotions: communicative emotions, situative emotions, and structural emotions. This notation allows a more precise description of emotions. Also, it allows defining possible, plausible relations between communicative emotions (cf. emotional expressions) and sequences of social signals to individual appraisal and regulation strategies. The latter can be triggered by elicited structural emotions, such as shame, which was our focus in this work.

On a conceptual level, the implications of MARSSI are twofold: 1) advancement of social signal classifiers with regard to an improved recognition of emotional aspects that can be related to structural emotions and 2) explanation of detected communicative emotions based on represented appraisal and regulation strategies and confidence values that are derived by the advanced social signal classifiers. The advancement of social signal classifiers is achieved by learning time and spatial relations of social signal sequences that are related to internal appraisal and regulation processes for a specific context. This process especially takes head and eye movements during communicative emotions into account reflecting the so far neglected aspect that human emotional expressions are directed. The MARSSI appraisal and regulation strategies allow possible explanations of detected communicative emotions concerning internal motivations. They are represented within the strategies and derived by related theories of emotion regulation.

We used a corpus-based approach to create our social signal classifiers in the context of job interviews. Some of the job interview questions are designed to elicit the structural emotion shame. Using MARSSI, we were able to model appraisal and regulation strategies that might occur in an applicant during a job interview. In a debriefing session, we used this knowledge together with our advanced social signal classifiers for analyzing each individual's social cues and for computing confidence values for modeled regulation strategies. An empathic virtual agent in the role of a job interview coach explains the regulation strategy with the highest confidence value. This enables the virtual coach to empathically address the possible elicited structural emotion shame explaining further details about the detected social cues.

MARSSI is a starting point for various types of research. The modeling of regulation strategies can be extended to cover other structural and even situational emotions. The notation of situational emotions could be exploited to learn how users emotionally remember a specific situation. An empathic agent might observe in the non-verbal behavior of users if past job interviews went bad. Since the advanced social signal classifiers rely on context information, we have to investigate if such classifiers can be applied in other contexts than the used job interview context. One important issue is the acceptance of such agents, especially if they can discuss their observations with the user. This could be exploited for agents to learn individual regulation patterns to refine the user model.

8.7 Acknowledgment

This work is partially funded by the German Ministry of Education and Research (BMBF) within the EmpaT project (funding code 16SV7229K). The authors thank the colleagues Thomas Anstadt, Ulrich Moser, and Eva Bänninger-Huber for providing their expertise in various phases of the presented work. The authors thank the Charamel GmbH and the TriCAT GmbH for realizing our requirements with regard to the virtual agents and the 3d environment.

9

Towards a Deeper Modeling of Emotions: The DEEP Method and its Application on Shame

Tanja Schneeberger[1], Mirella Hladký[1], Ann-Kristin Thurner[1], Jana Volkert[2], Alexander Heimerl[3], Tobias Baur[3], Elisabeth André[3], and Patrick Gebhard[1]

[1]German Research Center for Artificial Intelligence
[2]MSB Medical School Berlin
[3]Augsburg University

Published in Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction (ACII)
https://doi.org/10.1109/ACII52823.2021.9597446

This chaper contains the original version (with slight changes due to formatting) of our full paper presented at the 9th International Conference on Affective Computing and Intelligent Interaction (ACII 2021). During the online conference, the first author presented the paper orally to the conference attendees.

Abstract

Understanding emotions is key to Affective Computing. Emotion recognition focuses on the communicative component of emotions encoded in social signals. This view alone is insufficient for deeper understanding and computational representation of the internal, subjectively experienced component of emotions. This paper presents the DEEP method as a starting point for a deeper computational modeling of internal emotions. The method includes how to query individual internal emotional experiences, and it shows an approach to represent such information computationally. It combines social signals, verbalized introspection information, context information, and theory-driven knowledge. We apply the DEEP method exemplary on the emotion shame and present a schematic dynamic Bayesian network for modeling it.

Keywords: Emotion Modeling, Methods, Empirical Study

9.1 Introduction

Technological support for social and human affairs requires theories about the human psyche and societal structures. Within that context, the concept of emotion and the understanding of individuals' emotions seem very relevant. It comes with the hope and intention that through recognizing emotions, meaningful information about how an individual truly assesses or experiences a situation can be gathered (Tao & Tan, 2005). This information then could be exploited for a user model adapting to users' actual needs. The crux is that emotions – seen as individual internal experiences – cannot be recognized, at least with current approaches (Bradley & Lang, 2000; Feldman Barrett, 2017). Emotions have communicative components that are displayed in social signals and physiological parameters, but also internal components that reflect individual internal experiences (Feldman Barrett, 2017; Moser & von Zeppelin, 1996).

Research and applications of Affective Computing rely on understanding the emotions of the user. Thus, many attempts exist to infer user emotions exploiting different sources of data like speech, facial expressions, body gestures, and movement – unimodal as well as multimodal (Tao & Tan, 2005). This is still challenging, as emotions are a complex, hypothetical construct accompanied by changes in various components, including physiological reactions (e.g., heart rate) and behavioral components (e.g., facial expressions, gestures, and voice parameters; Feldman Barrett, 2017; Bradley & Lang, 2000).

Despite many efforts, a reliable assignment of observable reaction patterns to emotions, such as shame, fear, or surprise, remains unsolved (Feldman Barrett, 2017; Nathanson, 1994). This is not surprising knowing that emotions also have less directly accessible components, namely the subjective, internal experiences that might not be communicated (Feldman Barrett, 2017; Moser & von Zeppelin, 1996). Furthermore, in one particular situation, more than one emotion can arise (Harris, 1985), and emotion regulation allows masking internal emotional experiences (Gross, 2013).

However, for Affective Computing, it is crucial not only to understand and make assumptions based on the display of emotions but also to consider and represent connected individual experiences. Letting a system react solely based on the display of emotions might even be harmful, for example, when it is assumed that a smile is always mapped to a positive experience. However, a smile might be an expression of a negative affect like insecurity, shame, or fear. If such a system would react positively, it might even enforce the internal negative experience. This is relevant for social training systems, therapeutic assistance, and its acceptance and trustworthiness. Therefore, this paper presents the novel DEEP method with the purpose of gaining a deeper understanding of emotions that can be used for computationally understanding and modeling emotions. We show how we apply it to the emotion shame and how it could be realized with a dynamic Bayesian network.

9.2 Background and Related Work

9.2.1 Functions of Emotions

For understanding and modeling human emotions, especially in interactive social situations, understanding the functions of emotions is crucial. With their *intrapersonal functions*, emotions help us to operate quickly. As a rapid information-processing system, they enable acting with minimal thinking. They prepare the body for immediate action, for example, in dangerous situations, and are connected to perception, attention, inference, learning, goal choice, motivational priorities, physiological reactions, motor behaviors, and behavioral decision making (Tooby & Cosmides, 2008). The *interpersonal functions* of emotions refer to the role they play between two or more individuals. Humans express emotions verbally and nonverbally, which can be recognized by others (Elfenbein & Ambady, 2002). With their signal value, they influence interactions, for example, by evoking responses in interaction partners (Hwang & Matsumoto, 2019). Emotions also provide incentives for desired social behavior and therefore regulate social interactions (Keltner, 2003). The *socio-cultural functions* of emotions refer to the role they play in maintaining social order within a society. The cultural background defines which emotions are valued more (Tsai et al., 2006), how emotions are displayed and regulated (Hwang & Matsumoto, 2019). Humans manage, modify and express emotions through cultural display rules. These rules, usually learned in early childhood, define the appropriateness of emotional displays in certain social situations (Ekman & Friesen, 1969). As a result, especially negative emotions are often masked and not expressed openly (Nathanson, 1994).

9.2.2 Model of Emotions and Emotion Regulation

We follow a model of emotions that differentiates between internal (structural and situational) and external (communicative) components (Moser & von Zeppelin, 1996). *Structural components of emotions* represent information about the appraisal of one's attributes and actions. They are related to the self-image and provide information about its state. *Situational components of emotions* represent information linked to a topic or situation that has been experienced. *Communicative components of emotions* represent information communicated externally. They are verbally and non-verbally encoded in *sequences of social signals*, like vocal or facial expressions (Ekman, 1992). It also represents the information that is communicated to the person itself, like physiological reactions. Due to several processes, internal and external components might not match (Feldman Barrett, 2017; Moser & von Zeppelin, 1996). Influencing variables of the display of emotions are, for example, display rules (Ekman & Friesen, 1969) or emotion regulation processes (Gross, 2013).

Because most, if not all, emotions are regulated, understanding them is highly difficult. Emotion regulation refers to how humans try to influence which internal emotions they experience (Gross, 2013). This process can be conscious or unconscious. Emotion regulation can mean the regulation *by* emotions, referring to how emotions regulate something else, such as blood pressure, or it can mean the regulation *of* emotions, referring to how emotions themselves are regulated. People regulate emotions to avoid experiential and/or behavioral aspects of (negative) emotions such as anger, sadness, and shame. Gross specifies five types of regulation strategies: situation selection (choosing situations that are promising to experience wanted emotions or avoiding unpleasant situations), situational modification (modifying a given situation), attentional deployment (redirecting attention without changing a situation), cognitive change (changing one's appraisal of a situation in a way that alters the situation's emotional significance) and response modulation (influencing physiological, experiential, or behavioral responses). One form of response modulation is suppressing an emotional expression, like the effort to hide shame in an embarrassing situation (Gross, 2013).

9.2.3 Emotion Recognition and Emotion Modeling

There are many attempts to recognize human emotions in the field of Affective Computing (Picard et al., 2001; Soleymani et al., 2012; Valstar et al., 2016b), as well as to model them in computational emotion models (see Conati and Maclaren (2009) and Marsella et al. (2010) for an overview). Recently, interdisciplinary approaches are aiming to combine both (Belkaid & Sabouret, 2014; Gebhard et al., 2018).

The MARSSI model (Gebhard et al., 2018) relates appraisal rules and emotion regulation rules with social signal interpretation. It differentiates three functional dimensions of emotions: communicative, situative, and structural emotions. This notation allows a more accurate description of emotions. Also, it allows defining multiple possible, plausible relations between communicative emotions (cf. emotional expressions) and sequences of social signals to individual appraisal and regulation strategies. Elicited structural emotions can trigger the latter. However, this approach does not go beyond representing internal emotions as a label. Human internal emotions are always connected to subjective experiences and individual contexts that both can be computationally modeled (Sec. 9.3, 9.7).

For decades, researchers assumed that emotions have distinct patterns, like fingerprints, that are objectively observable (e.g., in facial expressions or brain activity). However, it seems that this is not the case. There is no one-to-one mapping between a specific set of facial muscle actions or vocal cues and any and every experience of emotion (Feldman Barrett, 2017). Moreover, the different measurements of emotions (physiological, behavioral, and experiential) are only feebly inter-correlated (Bradley & Lang, 2000). This might be why identifying objective, external means to measure the subjective, internal experience of emotions is complicated (Feldman Barrett, 2017). Therefore, other methods to acquire this information about emotions need to be explored.

One evident approach is to ask people about their subjective experience in self-reports (Feldman Barrett, 2004) in which they describe their internal experience (Izard et al., 1993; Watson et al., 1988). While questionnaires are suited for collecting quantitative data, for qualitative data, like internal experience, interviews and especially semi-structured interviews might be a more appropriate method (DiCicco-Bloom & Crabtree, 2006).

This work aims to develop a method to explore not only communicative components of emotions that are observable but also structural components of emotions that are internal.

9.3 Development of the Deep Method

All emotion recognition and emotion modeling methods can merely be seen as an approximation to individual internal experiences. We propose the Deep method, a multi-method approach to optimize this approximation combining four sources of information about one specific situation:

1. **Social signals**: Observation of communicated components of emotions that are encoded in social signals in the specific situation.

2. **Verbalized introspection information**: Self-reports that reflect a person's subjective experience gathered in semi-structured interviews after the specific situation with the aid of video material of the experienced situation.

3. **Context**: Social situation, display rules, roles of interaction partners in the specific situation, and information about the user like preferably applied regulation strategies, intention, personality, and others.

4. **Theory-driven knowledge**: Information about possible regulation strategies that can appear in social situations.

For a computational representation of this four sources of information, we anticipate a cognitive-oriented modeling with a dynamic Bayesian network (DBN) (Sec. 9.7).

9.3.1 Social Signals

Compared to approaches of emotion recognition, the DEEP method includes analyzing social signals communicated in a specific situation, too. For real-time analysis, we use the Social Signal Interpretation framework (SSI; Wagner et al., 2013) SSI especially allows synchronized processing of multiple sensor inputs in real-time. This includes the extraction of relevant features at runtime and the appliance of machine learning models, such as deep neural networks or support vector machines (SVM) for predicting single cues, such as changes in gaze or head direction, facial expressions, gestures, and postures.

9.3.2 Verbalized Introspection Information: Self-Report

The core of the DEEP method is the information from participants' self-reports about their internal experience. This information reflects the indirect introspection of the participant with the experience of interest lying in the past. It is recalled as a memory which is then observed and verbally described (Titchener, 1912). Semi-structured interviews obtain these self-reports suitably. They allow researchers to gain a deep understanding while following a guideline that ensures coverage of all important topics (Galletta, 2013), and the comparability of results (Polit & Beck, 2009). This interview form is versatile and flexible: It gives space for interviewees' individual reports and allows an exploration of the topic that may bring up yet unconsidered aspects. Moreover, it enables reciprocity between interviewer and interviewee (Galletta, 2013). For the use case of gaining information about internal experiences, semi-structured interviews are especially suited, as the collected data is rather personal, and retrieving it requires a careful and complex inquiry approach (Fylan, 2005).

To enhance the quality of verbalized introspection information from the semi-structured interviews, we propose to apply several techniques:

Supporting memory. The introspection follows immediately after the situation that is studied. To facilitate the process of remembering, experimenter and participant watch together a video of the studied situation (Galletta, 2013).

Creating comfort includes a positive atmosphere and the creation of a trustworthy relationship. The interviewer uses well-established nonverbal immediacy behaviors to show interest and engagement by orienting the body toward the interviewee, reducing interpersonal distance, smiling, showing open postures, and making eye contact (Imada & Hakel, 1977). On the verbal level, the interviewer self-discloses (Collins & Miller, 1994) and elicits an in-group feeling (Fu et al., 2012), for example, by confirming that it would also be difficult for him or her to talk about internal experiences. The set-up of the interview room ensures a feeling of privacy without disturbances. Interviewer and interviewee are seated at a 90° angle, optimal for interaction (Sommer, 1959).

Encouraging to speak openly is realized by showing interest and appreciation of what is said, for example, with verbal and non-verbal backchanneling signals (McNaughton et al., 2008). Psychotherapeutic questioning techniques encourage the interviewee to speak openly about every thought and feeling that comes to their mind (Will, 2006). Also, challenging questions are mixed with less stressful ones.

Reassuring information. To ensure a correct understanding of the interviewee's explanations, the interviewer paraphrases and summarizes the interviewee's answers after difficult questions. This facilitates the required interpretation of introspection results, as they are not self-explanatory (Titchener, 1912).

Selecting participants based on a priori formulated criteria is a valid method to improve qualitative research results (Galletta, 2013). The extent to which people can access their mental processes and states (e.g., emotions) varies inter-individually (Feldman Barrett, 2004; Fonagy et al., 2018). Hence, we pre-selected participants regarding *psychological mindedness* (Appelbaum, 1973) that has four factors: 1) the skill to discern connections between meanings and causes of behaviors, which requires both intact cognition, intuition, and empathy; 2) the goal of understanding the meaning of behaviors, which entails an interest in the way minds work; 3) self-directed psychological thinking; and 4) the "ability to engage in psychological thinking". However, selecting participants can affect the generalizability of the results (cf. Sec. 9.8).

9.3.3 Context

Emotions are generally elicited by (external or internal) stimulus events (Scherer, 2005). Information about this stimulus event and its context can improve modeling of an individual's internal experience. This context information may include knowledge about the interaction partners' cultures, as they highly influence how emotions are communicated (Ekman & Friesen, 1969). Moreover, it may include knowledge about the social situation and the roles of the interaction partners (Mesquita & Boiger, 2014). Context information can include knowledge about interaction partners' personal factors, such as preferred regulation strategies, psychological mindedness, mental load, intention, personality, as these can influence the internal experience in a specific situation.

9.3.4 Theory-driven knowledge

To understand, follow, and computationally represent individual situational experiences, a deeper knowledge of emotion and connected regulation processes is mandatory (Sec. 9.2.2).

9.4 Application of the Deep Method

The starting point for applying the Deep method is a previous study examining the emotion shame during the high-stakes situation of a job interview with a virtual interviewer (Schneeberger et al., 2019b). Results indicated that participants experienced shame in the shame-eliciting interview independent of the elicitor (human vs. virtual agent). They were based on observations of theoretically founded signals of shame and shame regulation. Self-reported questionnaire data regarding perceived discomfort in the shame-eliciting situation confirmed the finding.

However, as described before, analyzing the communicative component of emotions and self-assessment of emotions via questionnaires has several restrictions. Those especially apply for the emotion shame, as it leads to a highly unpleasant state that is difficult to cope with. Shame is rarely experienced consciously (Lewis, 2008; Moser & von Zeppelin, 1996). It is a social emotion and emerges particularly when individuals value the interaction partner's opinion (of them). The self fears rejection by the other in shameful situations (Hahn, 2001). Such a situation poses a threat to relationships and the self-concept by disclosing unfavorable information about the self. Thus, most often, it is immediately regulated unconsciously and not displayed openly (Gross, 2013; Moser & von Zeppelin, 1996; Nathanson, 1994). While they can manifest in observable behavior, shame experiences can remain solely internal, thus unobservable. Therefore, the observational and questionnaire data collected in the previous study may have not fully captured the very individual internal experience of shame. Also, shame is more challenging to talk about than other emotions (Keltner, 1996). Shame emerges when one notes failing to meet specific social standards. It is not elicited by a situation itself, but by an evaluation of that situation and oneself in it (Lewis, 2008). To conclude, the recognition of the highly complex emotion shame with existing methods might be impossible – thus requires a more careful and involved multi-method approach like the proposed Deep method.

When analyzing shame, regulation processes have to be taken into account. Nathanson describes four shame regulation strategies: 1) Withdrawal can manifest in avoiding eye contact and silence. The wish to hide or leave is characteristic of this strategy; 2) Attack Self is characterized by blaming oneself and addressing what others might accuse us of, thus regaining control. It can manifest in expressions of disgust or indignation toward oneself; 3) Avoidance is the effort to deceive oneself and others by pretending nothing has happened and directing the attention elsewhere; 4) Attack Other means answering a shame-triggering statement with a counterattack. Anger and disgust might be expressed towards the other. Here, termination of the relationship is accepted (Nathanson, 1994).

9.5 Study Methods

The present study's goal was to apply the DEEP method to a tested scenario. We oriented on our previous study examining shame (Schneeberger et al., 2019b). Approval was obtained from the project's ethical review board. Data was collected in November and December 2019. Additional material can be accessed at *osf.io/cv7t5*.

9.5.1 Screening and Participant Selection

In the study, participants were asked to elaborate on their internal experiences and possible explanations for emotions and cognition (Sec. 9.3). Therefore, we screened 35 psychology master students (28 female, $M_{\text{age}} = 23.97$ years, $SD_{\text{age}} = 2.20$ years) with the *Psychological Mindedness Scale* (Krupp et al., 2019). It consists of 34 items on four factors: interactive solution style, openness for change, access to one's feelings, willingness to try to understand oneself and others. Items were answered on a scale from 1 (*strongly disagree*) to 6 (*strongly agree*). Cronbach's Alpha ranged from .53 to .83. From the screened students, 27 reached a mean value beyond 4.5 (i.e., they either overall *agree* or *strongly agree* to be psychologically minded).

9.5.2 Participants

Due to the qualitative character of the study and the very detailed data analysis, we planned a sample size of $n = 10$. From the 27 invited participants, the first 10 (7 female) that registered participated in the study. Participants were aged between 22 and 32 years ($M = 24$, $SD = 3.06$) and had high values in the *Psychological Mindedness Scale* ($M = 4.87$, $SD = 0.97$). They were rewarded with 20€.

9.5.3 Procedure

Three days before the experiment, participants received the pre-questionnaire via email. On the interview day, participants were welcomed in the experimenter's room and informed about the procedure. After that, they filled in the shame experience questionnaire. Next, they were introduced to the job interview role-play for which they should imagine they applied for a student assistant position at their favorite university chair. They were told that a female virtual interviewer would conduct the interviews. Then, the experimenter guided them to the interviewer's office, which they entered alone. In the office, the virtual interviewer welcomed and asked them to sit down, then started the structured job interview conducted by the interactive social agent Susanne (Schneeberger et al., 2019b). The interview included two shame eliciting situations: "A brief question before we start. Where did you get this outfit? Somehow it doesn't really fit you." and "All the other applicants have already said what you said. You haven't exactly stood out.". During the job interview, the agent's turn-taking behavior was realized using a Wizard-of-Oz approach, with the wizard controlling when the agent starts talking. The experimental and technical set-up was like in (Schneeberger et al., 2019b). After the second shame-eliciting situation, the experimenter interrupted the interview, confirmed that it was planned like this, and handed them the shame experience questionnaire. The experimenter guided the participants back to the experimenter's room and revealed that the study's purpose was not the job interview itself but how they cope with the shame-eliciting situations. The post-interview followed. Afterward, participants answered the post interview assessment questionnaire. Finally, they were debriefed and paid. The whole procedure took ≈60 minutes.

9.5.4 Measurements

Demographics included age and gender and were covered in the pre-questionnaire.

Shame regulation strategies were measured with the Compass of Shame Scale (Elison et al., 2006). It assesses the use of the four shame-coping strategies described by Nathanson (1994): Withdrawal (WD), Attack Self (AS), Avoidance (AV), and Attack Other (AO). The questionnaire uses a description of a situation, for example, "When other people point out my faults" and reactions covering the four possible strategies: "I want to run away." (WD); "I feel like I can't do anything right." (AS); "I refuse to acknowledge those faults." (AV); "I point out their faults." (AO). In total 12 situations are described which results in 48 items. Each item was answered on a 5-point scale from 0 (*never*) to 4 (*almost always*). The questionnaire was translated into German and presented in the pre-questionnaire. Cronbach's Alpha ranged between .62 and .89.

Shame experience was measured before and after the job interview with six shame items from referring scales of the German version of the Differential

Emotion Scale (DES; Merten & Krause, 1993) and the Positive And Negative Affect Schedule (PANAS; Krohne et al., 1996). Two own items ("indignant" and "abashed") were added. To avoid priming, especially before the tasks, we included 11 shame-unrelated items of the DES as well as the PANAS. Items were answered on a scale ranging from 1 (*not at all*) to 5 (*very strong*). Due to increased Cronbach's Alpha, for the analysis, the item "shy" from the DES was removed. The resulting Cronbach's Alphas were .76 for the pre-test and .89 for the post-test.

Social signals in the shame eliciting-situations were observed and used for evaluating the occurance of shame and shame regulation as in (Schneeberger et al., 2019b).

The *post interview* took place after the two shame eliciting situations. It followed the guidelines described in Sec. 9.3.2. Participants were asked to talk openly about everything they think and feel, even if it seemed difficult. It was pointed out that the goal is to find out the very personal internal experience of the participant and that there is no right or wrong. Openness was also encouraged by emphasizing the research gains their reports bring. The interviewer asked if participants would like to see themselves during the two shame-eliciting situations on video. Consent was given by all, except one, participant. After the first situation the interviewer paused the video and asked the first broad question "What are your thoughts about this situation at the moment?". Further questions narrowed down the topic to internal experience, regulation strategies, bodily reactions, explanations for the emotions, cognition and behaviour, as well as connection between internal experience and social signals (e.g., smiling). Questions were formulated in a non-suggestive way so that participant's answers were genuine. The interview is designed that participants have the opportunity to mention feelings of shame on their own. If throughout the interview this did not happen, the interviewer explained that the job interview was supposed to elicit shame and provided a definition of shame. Then, participants were asked again about their internal experience in the situation. The procedure was repeated for the second situation.

Assessment of the post interview was measured with four self-constructed items on a scale from 1 (*strong disagreement*) to 5 (*strong agreement*). Items were "In the interview I openly said what I felt.", "It was difficult for me to talk about the experienced situation in the interview.", "The interview was agreeable.", "I was reluctant to talk about my feelings." (Cronbach's Alpha .93).

9.5.5 Post Interview Interpretation

The post-interviews were transcribed and jointly analyzed by three trained raters – one of them an experienced psychotherapist – regarding six variables: *1. Reaction in shameful situation.* We analyzed if a regulated shame vs. an open shame reaction is shown in the job interview and elaborated in the post interview. *2. Relationship.* As some shame regulation strategies are connected to a termination of the relationship with the other, we analyzed whether participants wish to maintain or terminate the relationship with the job interviewer. *3. Consciousness of shame in situation.* As shame is strongly unpleasant and poses a threat to the self-concept, it is often regulated and not consciously experienced. We analyzed whether participants were aware of shame in the shame-eliciting situation or not. *4. Mention of shame.* We analyzed whether participants mention on their own initiative that they felt shame. *5. Regulation strategies.* Based on the answers in the post interview, raters assessed which shame regulation strategies were applied. *6. Shame induction.* Based on elaborated shame regulation strategies and social signals in the shame-eliciting situation, raters assessed whether or not shame was elicited, also if shame was mentioned.

9.6 Study Results

In total, we video-recorded 20 shame-eliciting situations and audio-recorded 10 post-interviews.

9.6.1 Questionnaire Data

Shame experience. Participants reported significantly higher experienced shame after the job interview ($M = 1.90$, $SD = 0.80$) than before ($M = 1.18$, $SD = 0.24$), analyzed with a t-test for dependent measures ($t(9)$ = -2.66, p = .013, d = 0.85.

Shame regulation strategies. In the pre-questionnaire, participants self-reported their regulation strategies. In decreasing order, the regulation strategies were: Attack Self ($M = 2.18$, $SD = 0.66$); Withdrawal ($M = 1.96$, $SD = 0.49$); Avoidance ($M = 1.51$, $SD = 0.44$); Attack Other ($M = 1.25$, $SD = 0.36$).

Post Interview Assessment. Participants assessed their openness in the post interview and its agreeableness as high ($M = 4.30$, $SD = 0.87$).

9.6.2 Analyses of the Post Interview

The analysis of the situations and their respective elaborations during the post interview regarding *1. Reaction in shameful situation*, *2. Relationship*, *3. Consciousness of shame in situation*, *4. Mention of shame*, and *5. Regulation strategies* are enriched with quotes of participants (Table 9.1). In addition to Nathanson's regulation strategies (Nathanson, 1994), 15 other strategies were found, which are not elaborated in the present paper. Regarding *6. Shame induction*, raters assessed that in 18 situations, shame was induced. Shame induction was rated if the shame experience was mentioned explicitly or a shame regulation strategy

Table 9.1: Descriptive data and supporting quotes of participants.

Variable	Value	Frequency	Quote post interview
Shame reaction	not open	19	see variable **Regulation strategy**
	open	1	I showed shame rather open. Otherwise, I would not have apologized (#1)
Relationship	maintain	14	I still had the goal of getting the job (#3); I wanted to impress her (#4); Repair the image (#9)
	terminate	2	With people I don't like, I just don't care about them at all (#8); I would have liked to leave (#10)
	unclear	4	Go away (#4) vs. To still somehow impress her (#4)
Consciousness of shame	yes	13	I felt a bit inferior (#4); I think it was that sense of shame at that moment. This. Oh, I don't fit in here. What did I do wrong? (#6); I felt this unpleasant feeling consciously in the situation (#10)
	no	7	In the situation, I was angry. But now I realize that I tried to cover shame (#4)
Mention of shame	yes	15	I felt ashamed; It was unpleasant (#1); I felt unwell (#3); I felt personally attacked; I felt inferior (#4); This situation elicited more shame and discomfort (#5); I felt hurt (#10)
	no	5	Shame rather didn't get to me (#3)
Regulation strategy	Withdrawal	8	I avoided her gaze (#4, #5 #10); I had to let the situation pass by; I had to introspect and think about it (#9); I was overwhelmed, thus silent (#10)
	AttackSelf	1	I acknowledged she was right and I tried to improve my answer (#2)
	Avoidance	10	I diverted attention and did not really react on it (#2); I felt somehow offended and covered it up with the smile (#4); I thought it's funny (#7)
	AttackOther	6	I told her that she can either like my answer or not because I would not make up my professional expertise (#1); I smiled at her with a rather aggressive look (#4)
	Other	15	I wanted to prove her wrong (#2); I covered my shame with pride (#4); I accept myself how I am; I was self-confident (#6); It was due to the context, there wasn't much I could do (#8); If she attaches importance to something like that, then she is not a person I would attach importance to either (#9)

Note. In total, 20 shameful situations were analyzed. For the quotes, the participant number is given in parentheses.

was applied. The two remaining cases are unclear due to discrepancies between observed signals and information from the post interview.

9.6.3 Example Analysis

One study's goal was to examine the internal experience of participants throughout a shameful situation. Therefore, in a step-by-step process, we analyzed the video of the participant in the shame-eliciting situation as well as the verbalization of internal experience from the the post interview, connecting both data sources. Figure 9.1 presents the Deep analysis of the first shame-eliciting situation.

9.7 Conceptual Modeling Framework

Based on the theoretical foundation described in Section 9.3, we formulated a dynamic Bayesian network (DBN) modeling internal emotions (Figure 9.2). The DBN integrates the GenIE and SMILE library (*bayesfusion.com*) into our *open-source framework SSI* which enables updating the DBN with real-time observations from multiple channels (facial expressions, movements, voice etc), as well as external context information. In general, two types of nodes exist. Blue nodes represent information updated based on observations in the DBN, red nodes represent information inferred by the DBN. When it comes to understanding and recognizing emotions, considering various social signals is essential, for example, facial expression, gaze, upper body orientation. Those signals represent the observable result of the underlying regulated emotion and applied regulation strategy. The dashed lines represent temporal edges enabling the simulation and prediction of more complex motion sequences. The regulated emotion, and its corresponding manifestation in social signals, is a result of a regulation strategy that also manifests in social signals. However, it is also possible that individuals do not apply any regulation strategy at all. In turn, both are influenced by the individual context. This information includes knowledge about different aspects of the interaction, for example, cultural background, personality, or intention. Also, the individual context of a person primarily determines which internal emotion the person is experiencing. In this paper, we mainly focus on shame. However, the Deep method and the corresponding DBN can be applied to various internal emotions, like pride, admiration, and guilt. The internal emotion elicited by the individual context influences which regulation strategy a person is applying. The context, the regulation strategy, and the regulated emotion build the foundation for the verbalized introspection. This node represents a crucial aspect of the Deep method. Gathering information about how individuals experienced certain situations and why they reacted in a certain way ultimately helps to predict possible internal emotions reliably. The proposed network could be employed as an assistance system in psychotherapy or social training scenarios.

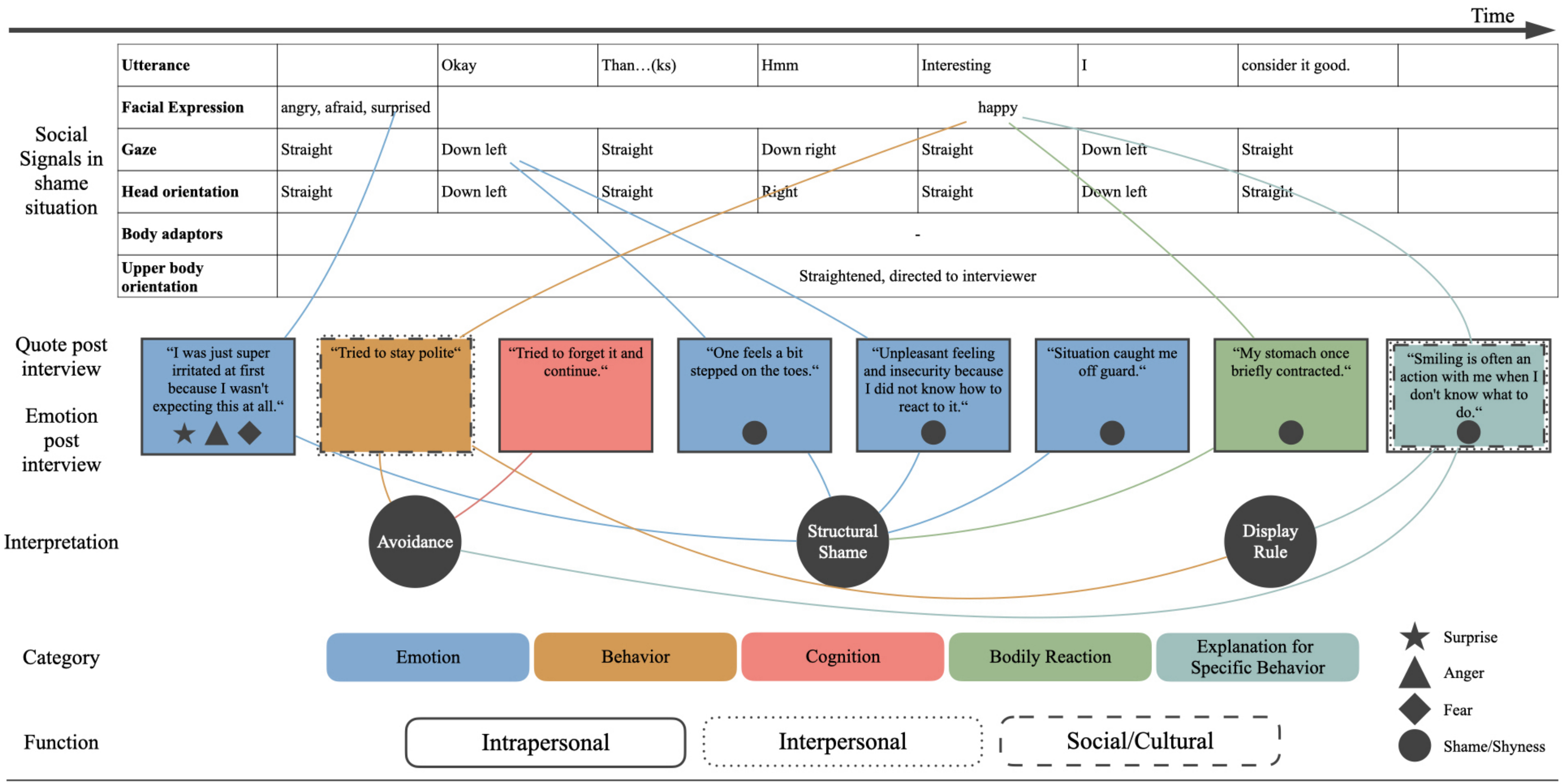


Figure 9.1: Analysis using the DEEP method for situation one ("A brief question before we start. Where did you get this outfit? Somehow it doesn't really fit you."), participant #3.

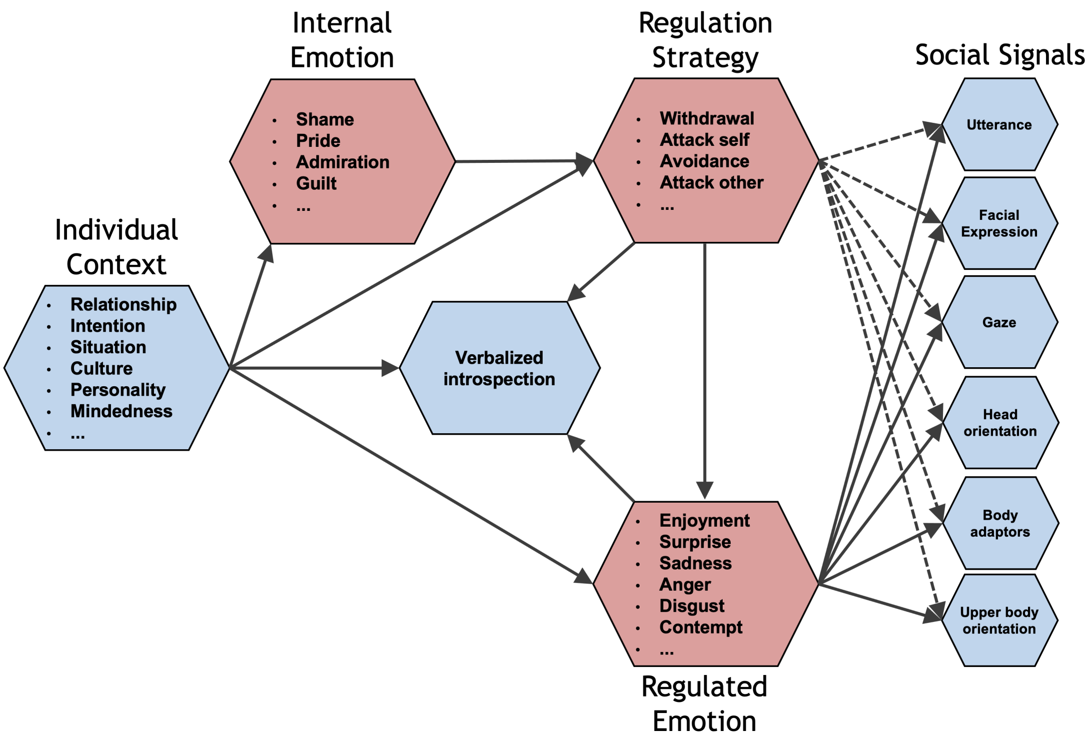


Figure 9.2: Schematic of a dynamic Bayesian network. Blue nodes (information updated based on observations); Red nodes (information inferred by the network); Solid edges (instantaneous causal effects), dashed edges (temporal causal effects).

9.8 Discussion

With this work, we introduce and apply a new multi-method approach to optimize the approximation of emotion understanding and modeling, the DEEP method. It combines four sources of information about one specific situation: social signals, context, self-reports, and theory-driven knowledge. We applied the introduced method to a situation in which participants experienced a shame elicitation in a job interview with a socially interactive agent. The questionnaire data indicates that the socially interactive agent can elicit shame in humans in the chosen situations. This replicates our previous study's finding showing once more that socially interactive agents can elicit an emotion of highly interpersonal nature (Schneeberger et al., 2019b). This questionnaire data is supported by observed social signals of shame and shame regulation as well as self-reports: Raters assessed that shame was successfully induced in 90% of the situations. When talking about the situation afterward, participants mentioned on their initiative that they had experienced shame for 75% of the situations. Though most of the situations induced shame, it was not displayed openly. Our results show that in all situations, except one, shame is not reflected in the communicative component of emotions by the ashamed person. This finding is consistent with previous work showing that shame is often not observable (Gross, 2013; Moser & von Zeppelin, 1996; Nathanson, 1994). Moreover, even though in most shame-eliciting situations, participants consciously felt shame, they wanted to maintain the relationship with the job interviewer. This demonstrates the crucial interpersonal function of shame, as it emerges when people note that they fail to meet social standards (Lewis, 2008).

Additionally to shame itself, the study examined shame regulation strategies. Results indicate a discrepancy between self-assessment of regulation strategies usage in questionnaires and their observed occurrence. In self-assessment, the regulation strategy Attack Self was most commonly reported, whereas it was least applied in observed situations. This might indicate that context plays a more important role than people's general tendency to apply a specific shame regulation strategy.

The introduced Deep method is a starting point for future research as it enables an understanding of internal emotions. Participants successfully verbalized their internal experience and confirmed that they spoke openly during the post interview. With minor adaptions, it is possible to apply the method to other emotions that are often regulated. Except the interpretation level (Figure 9.1), the analysis is applicable without little adaptations to other emotions in other situations. Also, the post-interview questions were on a general level regarding internal experience, regulation strategies, bodily reactions, explanations for emotions, cognition, and behavior, as well as the connection between internal experience and social signals. However, possible emotion regulation strategies need to be adapted for the interpretation of the post interview answers as well as in the DBN.

9.9 Limitations and Future Work

The Deep Method is costly as it involves time-consuming data collection and analysis with trained interviewers and raters. Therefore, we are working on a questionnaire covering the semi-structured interview. As a method, a questionnaire could reduce the threshold of participants talking about a difficult emotional experience (Keltner, 1996). Moreover, applying automatic analysis tools could be an option to make the analysis process more efficient and more standardized.

For now, due to the time-consuming data analysis this method involves, its application is limited to one emotion and a small, pre-selected sample. If and how this procedure can be applied to individuals with a lower Psychological Mindedness score is a topic of future research. Theoretically, a lower score implies poorer ability to access and verbalize internal experiences, which results in limitations in perceiving, differentiating, or naming affect (Appelbaum, 1973).

A major challenge for computational representation of internal emotional experience is expanding all group nodes (e.g., Figure 9.2, Individual Context, Regulation Strategy) into basal distinguished concept nodes. This requires an even deeper understanding of how concepts (e.g., culture, relationship, personality) interfere. More in reach is a more fine-grained presentation of internal states like appraisals, goals, motivations, and taking user actions as input. Existing DBN models (e.g., Conati & Maclaren, 2009) could be combined with the DEEP network schema.

9.10 Conclusion

This work presents the DEEP method of how to query individual internal emotional experiences and shows an approach how to represent such information computationally. Similar to other methods, the presented method is an approximation because measuring the "actual" emotional experience is daunting, as – by nature – it is internal and subjective. However, this work presents a first important step towards a deeper understanding and modeling of emotions as internal, highly subjective experiences that are mostly not openly displayed. The emotion analysis with the DEEP method includes social signals, context, self-reports, and theory-driven knowledge.

9.11 Acknowledgment

We thank Svenja Taubner and Eva Bänninger-Huber for their advise in the DEEP project. We thank Saarland Informatics Campus for the provision of the experiment rooms.

10
General Discussion

The aim of this dissertation was to conduct studies following the Computers-Are-Social-Actors paradigm examining social behavior towards SIAs and to computationally model the observed human social behavior towards SIAs. Therefore, it presents a line of work that investigates how humans affectively react towards machines that have a SIA as an interface. It also introduces two approaches to computationally modeling affective responses, taking into account emotion regulation and emotion as an individual experience. In summary, this dissertation provides evidence in support of the ethopoeia approach, but also shows the complexity of modeling affect in the computer. In what follows, the six contributions of this thesis are summarized. Afterwards, the general strengths and weaknesses of this dissertation will be discussed, followed by a section on implications for research practice and directions for future research.

The first contribution is the presentation of PARLEY, an interactive system to train difficult social situations that was used in an adapted form for all studies conducted for this thesis. During the course of the work for this thesis, the PARLEY system was greatly extended and refined. The system provides a means of interaction for users to practice social situations of all kinds in a protected space. The interaction with the SIAs is realized as natural as possible compared to human-human interaction. Therefore, the PARLEY-SIAs exhibit credible social communication behavior both while speaking and listening. In general, they have a wide repertoire of facial expressions and gestures, which makes it possible to create realistic emotional expressions. In addition, they can mirror specific behavior of the user, such as smiling or nodding, to create the illusion of mimicry. The degree of mirroring in terms of frequency and time delay can be defined according to the needs of the situation being practiced. During conversation, the SIAs are designed to be interrupted by the user. The SIAs suspend the interaction, handle the interruption, and then reconstruct and resume the initial interaction. This highly reactive and adaptive ability makes the interaction more natural than the state-of-the-art SIAs and even speech interfaces like Siri. When listening, the SIAs show backchanneling behavior, giving the user a sense of being heard and understood. On the user side, the system is designed to require no explicit input devices, such as keyboards. The system is further enhanced by a remote control, the Study Master. With the Study Master, the experimenter can control the behavior of the SIA in real time during an experiment, for example, depending on an answer given by the participant. This so-called Wizard-of-Oz approach is applicable not only in a laboratory with the physical presence of both participant and experimenter, but also as a service over the internet. A participant can interact with a SIA controlled by an experimenter, and both can be located anywhere. Finally, the system is implemented in a way that allows researchers with a non-technical background to realize interactions with SIAs.

The second contribution is a study that examines whether SIAs can elicit the social emotion shame as humans do. Following the classic CASA paradigm, a SIA in the role of a job interviewer was compared to a human role-player. For the shame elicitation, five pre-tested situations were embedded in the job interview, reflecting different associations to the self that might elicit shame. The analysis of the observational data as well as the self-assessment questionnaires showed that the SIA attacking the self of participants was able to elicit the same level of shame as the human. This finding was insofar remarkable as it revealed that SIAs could take on a considerable role for a human user by making the human feel dependent and fear the reaction of the SIA.

The third contribution of the thesis showed that SIAs could adopt roles with a certain degree of authority. Following the classic CASA paradigm, a SIA in the role of an instructor was compared to a human role-player. In the experiment, a cover story of a creativity test was used to instruct participants to obey to a maximum amount of 18 tasks, increasing in stress and shame levels. The amount of obedience was measured with a behavioral variable, namely which task the participants refused to obey. Perceived instructor authority was measured with a questionnaire. The affective reaction was measured with self-assessment questionnaires asking for stress and shame. The results indicated that the SIA had the same authority as the human instructor and that it was also able to elicit the same level of the negative feelings stress and shame. Overall, the study provided further evidence for the validity of the Media Equation, as SIAs appear to be able to influence humans even when it comes to tasks that are uneasy to perform. Furthermore, it replicates the finding of the second contribution, showing that shame can be elicited to the same extent by SIAs and humans.

The fourth contribution is a stress management training using biofeedback derived from the cardiovascular response of the heart rate variability with a SIA as a biofeedback trainer. The training was evaluated with a subject-matter expert interview and an experiment comparing the novel approach with a stress management training using stress diaries. The training was assessed regarding its effectiveness in teaching stress management strategies for stressful social situations. The experiment results indicated that the novel approach reduced the self-assessed stress levels immediately after the training, as well as in a socially stressful task. Moreover, participants who received the training with the SIA rated their performance in the socially stressful task higher than participants who received stress diaries. Taken together, it appears that the virtual stress management training with a SIA as a trainer is a valid method for learning techniques on how to cope with stressful situations. Compared to the experiments in the second and third contribution, in this experiment we did not compare a SIA with a human trainer. Therefore, we cannot conclude whether the effectiveness of the stress management training with a SIA is similar to one with a human trainer.

With the fifth contribution, MARSSI was introduced. This computational model of user emotions for SIAs combines a simulation of appraisal and regulation processes with a social signal interpretation. The model was evaluated on the social emotion shame using the corpus collected in the shame-eliciting study (Chapter 5). It was the first computational model of emotions, considering that communicative, emotional expressions are not necessarily directly related to internal emotional states.

The sixth contribution is the DEEP method, a method for gaining a deeper understanding of emotions that can be used for computational understanding and modeling emotions. The method describes how to query individual internal emotional experiences and shows an approach to represent such information computationally. Social signals, verbalized introspection information, context information, and theory-driven knowledge are combined to approximate the user's affect. The DEEP method was applied exemplarily to the emotion shame and presented a schematic dynamic Bayesian network for modeling it. It was the first method in Affective Computing going beyond emotion recognition towards a deeper understanding and modeling of emotions as internal, highly subjective experiences that are mostly not openly displayed.

Overall, the work done for this thesis supports the "ethopoeia" approach in the media equation assumption (Nass & Moon, 2000; Reeves & Nass, 1996). It provides further evidence that humans react socially towards machines with a SIA interface. These results confirm once again that humans automatically and unconsciously apply social rules to their interactions with computers. Even more, when interacting with SIAs, humans behave in their inherently social way, showing affective reactions that were previously studied only in human-human interactions. These new findings go beyond previous results showing that humans do automatically apply social rules to their interactions with SIAs, because humans are inherently social - independently of their interaction partner (Deladisma et al., 2007; Gratch et al., 2007; Hoffmann et al., 2009; Kopp et al., 2005; Krämer et al., 2013; Krämer et al., 2018; Lucas et al., 2014; Sproull et al., 1996; Weitz et al., 2019, 2021). Together with existing evidence from experiments done with SIAs within the CASA paradigm, the work for this thesis shows that SIAs can represent a human object and thus take on crucial roles in the lives of human users. This is especially the case when they are equipped with a computational emotional model that simulates the user's emotions. Therefore, the work done for this thesis also presents two approaches for modeling the affective reactions of users shown in interactions with SIAs. Both approaches show the complexity of modeling social emotions, which has not been done before (Marsella & Gratch, 2014).

10.1 Limitations

In addition to the limitations outlined in the respective chapters, there are more general limitations that need to be addressed.

First, the studies presented for this dissertation following the CASA paradigm have the same limitation as all studies following this paradigm. Although great care was taken to minimize the effect of the human experimenter, for example, by being alone in a room with the SIA during the experiment, it is still possible that the imagined presence of others may have influenced the participants' social responses. Of course, the participants were informed in advance that audio and video recordings would be made. Therefore, it is possible that they imagined the people who would later see and hear these recordings when interacting with the SIA. Overcoming this limitation is a challenge in scientific studies, as observing participants is at the core of these studies. However, future developments of SIAs that could be used for personal use may provide insight into reactions towards them in a more naturalistic setting.

In addition, the participants in the experiments did not experience real-life situations, and it remains to be shown whether the results can be generalized to real-life situations. Regarding the job interview (Chapters 5, 8, & 9), it was a mock interview for a hypothetical research assistant position. Applying the research design to a real application situation seems highly unethical. It may be that people would feel even more ashamed in a real interview because the stakes are higher. In the study examining obedience towards SIAs, participants thought they were evaluating a novel creativity test. Whether people also obey SIAs in a real-life situation, such as when SIAs act as security guards giving instructions in an evacuation, is an interesting research question. However, investigating this in a study may raise ethical concerns because it would require participants to believe that they were in a threatening situation. The socially stressful task used to evaluate the biofeedback training was also an experimental situation. It seems feasible to investigate whether the stress management training also affects real-life stressful situations. In fact, it is planned to evaluate the biofeedback stress management training with an SIA as trainer in a naturalistic setting in a vocational training center.

Third, as with most studies, the study samples limit the generalizability of the findings. The participants in the studies conducted for this dissertation were all highly educated, mostly Caucasian, female, and in their early twenties. Therefore, the results may not apply to a more diverse group of users. In particular, the results may be significantly different for people who do not use technology on a regular basis.

10.2 Future Research and Implications

The application of SIAs in research is substantial and growing. The areas of their application are vast. They range from assistive and health technologies to education and computer games. Both will lead to their increased presence in everyday life – at least in parts of the world. Perhaps the most significant growth potential lies in the conversational capabilities of SIAs. Enhanced conversational AI, more specifically machine learning (ML) and deep learning (DL) models, natural language understanding and processing techniques, and dialog management systems to understand user input and generate natural language responses (Ruane et al., 2019) will greatly improve the interaction with SIAs. Especially if user affect is also taken into account, enhanced conversational AI could allow SIAs to formulate their statements based on simulated user affect. Future studies will be able to examine whether these developments will affect affective reactions towards SIAs. It could be expected that social reactions to SIAs in general will become more pronounced as the social capabilities of SIAs are enhanced (Krämer & Manzeschke, 2021).

Also, research examining the interaction between humans and SIAs would profit from field experiments to increase external validity. To date, most studies conducted with SIAs have been experimental laboratory studies, including the studies conducted for this dissertation. Experimental laboratory studies typically have the advantage of ensuring internal validity. However, to study interactions between humans and SIAs in everyday encounters, field studies need to be conducted (Krämer & Manzeschke, 2021).

Another very interesting question is how affective reactions towards SIAs develop during long-term interactions. Research with this focus goes beyond interaction and examines what kinds of relationships people form with SIAs (Kory-Westlund et al., 2022). In human-human relationships, higher quality relationships have been found to lead to better outcomes, such as relationships between therapists and patients (Wampold, 2015) or coaches and clients (Frates et al., 2011). Although long-term social-emotional relationships with users were introduced in 2003 (Bickmore, 2003), there are still less than 25 scientific publications reporting studies with more than four interactions (Krämer & Manzeschke, 2021). In these studies, SIAs mostly attempt to motivate participants to engage in healthy exercise behaviors, such as establishing a healthy walking routine (Bickmore et al., 2005b; Bickmore et al., 2013a). Although it is daunting for researchers to study long-term interactions between humans and SIAs over weeks, months, or years, more research in this direction is needed to understand how human-SIA relationships develop. This is a critical factor in learning the value that SIAs can bring to personal lives.

Moreover, it still remains unclear why some people react more affectively to SIAs and others less so. In the studies presented in the previous chapters, some participants were highly engaged in the interactions with the SIAs, while others were not. In general, most people seem to connect with SIAs, as many studies report the positive effects of SIAs. However, in the studies reported for this thesis, there were a number of participants who were not much affected by the SIAs. Therefore, future work could examine the reasons and determinants behind people's willingness to emotionally connect with an SIA.

The application of SIAs is a rapidly developing area not only in research but also in consumer-facing commercial applications, for example in the health context (Bickmore, 2022) and in games (Prada & Rato, 2022). As in other areas, companies are developing products that have not been scientifically evaluated. While this may be negligible in the context of gaming, it raises ethical issues in the context of health. To protect people from AI systems, including applications with SIAs, the European Commission published a proposal for a law on Artificial Intelligence in April 2021, the AI Act[1]. The proposal sets out a methodology for classifying systems using artificial intelligence into three risk categories: unacceptable risk, high-risk, and non-high risk. It also sets out mandatory requirements for trustworthy AI and assessment procedures for high-risk applications before these systems can be placed on the Union market. It will be up to the members of the European Council and the European Parliament to decide whether the AI Act will become law.

10.3 General Conclusion

> "Everyone is much more simply human than otherwise."
>
> - Harry Stack Sullivan, 1953 -

This dissertation is one piece of interdisciplinary research that seeks to understand human affective reactions towards computers and to represent these reactions in a computer. It joins a series of studies with a 30-year tradition of comparing human and technical interaction partners. Overall, there seems to be evidence that humans react very humanely to SIAs, and that it is possible to model this affective reaction in a computer. However, it remains to be seen how interactions between humans and SIAs will look in everyday life and whether reactions towards them will be shaped by time. Therefore, more interdisciplinary collaborations in computer science and psychology are needed to conduct research that shapes the interaction between humans and SIAs.

[1] artificialintelligenceact.eu/the-act/

Bibliography

Abe, J. A. (2004). Shame, guilt, and personality judgment. *Journal of Research in Personality*, *38*(2), 85–104. https://doi.org/10.1016/S0092-6566(03)00055-2

Agrawal, S., & Williams, M.-A. (2017). Robot authority and human obedience: A study of human behaviour using a robot security guard. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 57–58. https://doi.org/10.1145/3029798.3038387

Alford, W. K., Malouff, J. M., & Osland, K. S. (2005). Written emotional expression as a coping method in child protective services officers. *International Journal of Stress Management*, *12*(2), 177–187. https://doi.org/10.1037/1072-5245.12.2.177

Ali, M. R., Razavi, S. Z., Langevin, R., Al Mamun, A., Kane, B., Rawassizadeh, R., Schubert, L. K., & Hoque, M. E. (2020). A virtual conversational agent for teens with autism spectrum disorder: Experimental results and design lessons. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 1–8. https://doi.org/10.1145/3383652.3423900

Ali, M. R., Van Orden, K., Parkhurst, K., Liu, S., Nguyen, V.-D., Duberstein, P., & Hoque, M. E. (2018). Aging and engaging: A social conversational skills training program for older adults. *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, 55–66. https://doi.org/10.1145/3172944.3172958

Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssafidou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., Jones, H., Ochs, M., Pelachaud, C., Porayska-Pomsta, K., Rizzo, P., & Sabouret, N. (2013). The TARDIS framework: Intelligent virtual agents for social coaching in job interviews. *Proceedings of the 10th International Conference on Advances in Computer Entertainment*, 476–491.

Andrade, A. D., Bagri, A., Zaw, K., Roos, B. A., & Ruiz, J. G. (2010). Avatar-mediated training in the delivery of bad news in a virtual world. *Journal of Palliative Medicine*, *13*(12), 1415–1419. https://doi.org/10.1089/jpm.2010.0108

App, B., McIntosh, D. N., Reed, C. L., & Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: How may depend on why. *Emotion*, *11*(3), 603–617. https://doi.org/10.1037/a0023164

Appel, M., Lugrin, B., Kühle, M., & Heindl, C. (2021). The emotional robotic storyteller: On the influence of affect congruency on narrative transportation, robot perception, and persuasion. *Computers in Human Behavior*, *120*, Article 106749. https://doi.org/10.1016/j.chb.2021.106749

Appelbaum, A., Stephen. (1973). Psychological-mindedness: Word, concept and essence. *International Journal of Psycho-Analysis*, *54*(1), 35–46.

Aylett, R., Vannini, N., Andre, E., Paiva, A., Enz, S., & Hall, L. (2009). But that was in another country: Agents and intercultural empathy. *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, 329–336.

Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*(10), 814–819. https://doi.org/10.1111/j.1467-9280.2005.01619.x

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, *17*(1), 102–141. https://doi.org/10.1080/10508400701793141

Bainbridge, W. S. (2004). *Berkshire encyclopedia of human-computer interaction: Volume 2*. Berkshire Publishing Group LLC.

Balcar, K. (2011). Trends in studying emotions. In R. Trnka, K. Balcar, & M. Kuška (Eds.), *Re-constructing emotional spaces: From experience to regulation* (pp. 1–32). College of Psychosocial Press.

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). Openface: An open source facial behavior analysis toolkit. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 1–10. https://doi.org/http://dx.doi.org/10.1109/WACV.2016.7477553

Bänninger-Huber, E. (1996). *Mimik Übertragung Interaktion: Die Untersuchung Affektiver Prozesse in der Psychotherapie*. Huber.

Bänninger-Huber, E., Moser, U., & Steiner, F. (1990). Mikroanalytische Untersuchung affektiver Regulierungsprozesse in Paar-Interaktionen. *Zeitschrift für Klinische Psychologie*, *19*(2), 123–143.

Bänninger-Huber, E., & Steiner, F. (1992). Identifying microsequences: A new methodological approach to the analysis of affective regulatory processes. In M. Leuzinger-Bohleber (Ed.), *"Two Butterflies on My Head..."* (pp. 257–276). Springer.

Bänziger, T., Tran, V., & Scherer, K. R. (2005). *The Geneva emotion wheel: A tool for the verbal report of emotional reactions* [poster presentation].

Conference of the International Society for Research on Emotion, Bari, Italy.

Bao, A.-M., Meynen, G., & Swaab, D. (2008). The stress system in depression and neurodegeneration: Focus on the human hypothalamus. *Brain Research Reviews*, *57*(2), 531–553. https://doi.org/10.1016/j.brainresrev.2007.04.005

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, *94*(6), 1394–1411. https://doi.org/10.1037/a0016532

Bartneck, C., Bleeker, T., Bun, J., Fens, P., & Riet, L. (2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn*, *1*(2), 109–115. https://doi.org/10.2478/s13230-010-0011-3

Bates, J. E. (1975). Effects of a child's imitation versus nonimitation on adults' verbal and nonverbal positivity. *Journal of Personality and Social Psychology*, *31*(5), 840–851. https://doi.org/10.1037/h0076677

Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., & Scherer, S. (2013). Cicero-towards a multimodal virtual audience platform for public speaking training. *Proceedings of the 13th ACM International Conference on Intelligent Virtual Agents*, 116–128. https://doi.org/10.1007/978-3-642-40415-3_10

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. In R. Zukauskiene (Ed.), *Interpersonal development* (pp. 57–89). Routledge.

Baur, T., Mehlmann, G., Damian, I., Lingenfelser, F., Wagner, J., Lugrin, B., André, E., & Gebhard, P. (2015). Context-aware automated analysis and annotation of social human–agent interactions. *ACM Transactions on Interactive Intelligent Systems*, *5*(2), 1–33. https://doi.org/10.1145/2764921

Bavelas, J. B., & Chovil, N. (1997). Faces in dialogue. In J. A. Russell & J. M. Fernadez-Dols (Eds.), *The psychology of facial expression* (pp. 334–346). Cambridge University Press.

Bedwell, W. L., & Salas, E. (2010). Computer-based training: Capitalizing on lessons learned. *International Journal of Training and Development*, *14*(3), 239–249. https://doi.org/10.1111/j.1468-2419.2010.00355.x

Bee, N., André, E., Elisabeth, Vogt, T., & Gebhard, P. (2010). Close engagements with artificial companions: Key, social, psychological, ethical and design issues. In Y. Wilks (Ed.), *Natural language processing* (pp. 131–142). John Benjamins Publishing.

Bègue, L., Beauvois, J.-L., Courbet, D., Oberlé, D., Lepage, J., & Duke, A. A. (2015). Personality predicts obedience in a Milgram paradigm. *Journal of Personality*, *83*(3), 299–306. https://doi.org/10.1111/jopy.12104

Belkaid, M., & Sabouret, N. (2014). *A logical model of Theory of Mind for virtual agents in the context of job interview simulation*. https://doi.org/10.48550/ARXIV.1402.5043

Benecke, C. (2002). *Mimischer Affektausdruck und Sprachinhalt* (Doctoral dissertation). Saarland University.

Berkelaar, B. L., & Buzzanell, P. M. (2014). Cybervetting, person-environment fit, and personnel selection: Employers' surveillance and sensemaking of job applicants' online information. *Journal of Applied Communication Research*, *42*(4), 456–476. https://doi.org/10.1080/00909882.2014.954595

Bernardi, L., Porta, C., Gabutti, A., Spicuzza, L., & Sleight, P. (2001). Modulatory effects of respiration. *Autonomic Neuroscience*, *90*(1-2), 47–56. https://doi.org/10.1016/S1566-0702(01)00267-3

Bernardini, S., Porayska-Pomsta, K., & Sampath, H. (2013). Designing an intelligent virtual agent for social communication in autism. *Proceedings of the 9th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 9–15. https://doi.org/10.1609/aiide.v9i1.12688

Bickmore, T. W. (2003). *Relational agents: Effecting change through human-computer relationships* (Doctoral dissertation). Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/36109

Bickmore, T. W. (2022). Health-related applications of socially interactive agents. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 2: Interactivity, platforms, application* (pp. 403–436). Association for Computing Machinery. https://doi.org/10.1145/3563659.3563672

Bickmore, T. W., Caruso, L., Clough-Gorr, K., & Heeren, T. (2005a). 'It's just like you talk to a friend' relational agents for older adults. *Interacting with Computers*, *17*(6), 711–735. https://doi.org/10.1016/j.intcom.2005.09.002

Bickmore, T. W., Gruber, A., & Picard, R. W. (2005b). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, *59*(1), 21–30. https://doi.org/10.1016/j.pec.2004.09.008

Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., & O'Donnell, J. (2010). Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers*, *22*(4), 289–298. https://doi.org/10.1016/j.intcom.2009.12.001

Bickmore, T. W., Pfeifer, L., & Schulman, D. (2011). Relational agents improve engagement and learning in science museum visitors. *Proceedings of the 11th ACM International Conference on Intelligent Virtual Agents*, 55–67. https://doi.org/10.1007/978-3-642-23974-8_7

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *Transactions on Computer-Human Interaction*, *12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bickmore, T. W., Rubin, A., & Simon, S. R. (2020). Substance use screening using virtual agents: Towards automated screening, brief intervention, and referral to treatment (SBIRT). *Proceedings of the 20th International Conference on Intelligent Virtual Agents*, 1–7. https://doi.org/10.1145/3383652.3423869

Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L., & Paasche-Orlow, M. K. (2013a). A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society*, *61*(10), 1676–1683. https://doi.org/10.1111/jgs.12449

Bickmore, T. W., Vardoulakis, L. M. P., & Schulman, D. (2013b). Tinker: A relational agent museum guide. *Autonomous Agents and Multi-Agent Systems*, *27*(2), 254–276. https://doi.org/10.1007/s10458-012-9216-7

Blouin, N., Deaton, J., Richard, E., & Buza, P. (2014). Effects of stress on perceived performance of collegiate aviators. *Aviation Psychology and Applied Human Factors*, *4*(1), 40–49. https://doi.org/10.1027/2192-0923/a000054

Bosma, W., & André, E. (2004). Exploiting emotions to disambiguate dialogue acts. *Proceedings of the 9th International Conference on Intelligent User Interfaces*, 85–92. https://doi.org/10.1145/964442.964459

Bosse, T., Gerritsen, C., & de Man, J. (2016). An intelligent system for aggression de-escalation training. *Proceedings of the 22nd European Conference on Artificial Intelligence*, 1805–1811. https://doi.org/10.3233/978-1-61499-672-9-1805

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology*, *77*(3), 343–352. https://doi.org/10.1016/j.biopsycho.2007.11.008

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 242–276). Oxford University Press.

Brown, B. B. (1977). *Stress and the art of biofeedback*. Harper & Row.

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology* (2nd ed.). John Wiley & Sons Inc.

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, *64*(1), 1–11. https://doi.org/10.1037/a0010932

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human-computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior*, *16*(6), 553–574. https://doi.org/10.1016/S0747-5632(00)00029-7

Burke, S. L., Bresnahan, T., Li, T., Epnere, K., Rizzo, A., Partin, M., Ahlness, R. M., & Trimmer, M. (2018). Using virtual interactive training agents (ViTA) with adults with autism and other developmental disabilities. *Jour-*

nal of Autism and Developmental Disorders, *48*(3), 905–912. https://doi.org/10.1007/s10803-017-3374-z

Buss, A. H. (1980). *Self-consciousness and social anxiety*. Freeman.

Cafaro, A., Glas, N., & Pelachaud, C. (2016). The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. *Proceedings of the 2016 International International Conference on Autonomous Agents and Multiagent Systems*, 911–920.

Cameron, J. (2009). *Avatar*. 20th Century Fox.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*(3), 655–702. https://doi.org/10.1111/j.1744-6570.1997.tb00709.x

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*(2), 205–218. https://doi.org/10.1037/0022-3514.70.2.205

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, *56*(2), 267–283. https://doi.org/10.1037/0022-3514.56.2.267

Cassell, J., Sullivan, J., Churchill, E., & Prevost, S. (2000). *Embodied conversational agents*. MIT Press.

Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., & Pecchia, L. (2015). Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control*, *18*, 370–377. https://doi.org/10.1016/j.bspc.2015.02.012

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 334–361).

Cheok, A. D., & Levy, D. (2018). *Love and Sex with Robots: Third International Conference*.

Chiesa, A., & Serretti, A. (2009). Mindfulness-based stress reduction for stress management in healthy people: A review and meta-analysis. *The Journal of Alternative and Complementary Medicine*, *15*(5), 593–600. https://doi.org/10.1089/acm.2008.0495

Chittaro, L., & Sioni, R. (2014). Affective computing vs. affective placebo: Study of a biofeedback-controlled game for relaxation training. *International Journal of Human-Computer Studies*, *72*(8-9), 663–673. https://doi.org/10.1016/j.ijhcs.2014.01.007

Chollet, M., Ghate, P., & Scherer, S. (2018). A generic platform for training social skills with adaptative virtual agents. *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, 1800–1802.

Chollet, M., Sratou, G., Shapiro, A., Morency, L.-P., & Scherer, S. (2014). An interactive virtual audience platform for public speaking training. *Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems*, 1657–1658.

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S. (2015). Exploring feedback strategies to improve public speaking: An interactive virtual audience framework. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1143–1154. https://doi.org/10.1145/2750858.2806060

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242. https://doi.org/10.1006/cogp.1998.0693

Cohen, S. (1994). *Perceived Stress Scale*. Mindgarden.

Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological stress and disease. *JAMA*, *298*(14), 1685–1687. https://doi.org/10.1001/jama.298.14.1685

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, *116*(3), 457–475. https://doi.org/10.1037/0033-2909.116.3.457

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*(7-8), 555–575. https://doi.org/10.1080/08839510290030390

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267–303. https://doi.org/10.1007/s11257-009-9062-8

Cooper, S., Di Fava, A., Vivas, C., Marchionni, L., & Ferro, F. (2020). Ari: The social assistive robot and companion. *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication*, 745–751. https://doi.org/10.1109/RO-MAN47096.2020.9223470

Cormier, D., Newman, G., Nakane, M., Young, J. E., & Durocher, S. (2013). Would you do as a robot commands? An obedience study for human-robot interaction. *Proceedings of the 1st International Conference on Human-Agent Interaction*, 1–8.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, *24*(2), 227–245. https://doi.org/10.1177/0894439305281503

Damasio, A. (2017). *Im Anfang war das Gefühl: Der biologische Ursprung menschlicher Kultur*. Siedler.

Damian, I., Baur, T., Lugrin, B., Gebhard, P., Mehlmann, G., & André, E. (2015a). Games are better than books: In-situ comparison of an interactive job interview game with conventional training. *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, 84–94. https://doi.org/10.1007/978-3-319-19773-9_9

Damian, I., Tan, C. S., Baur, T., Schöning, J., Luyten, K., & André, E. (2015b). Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 565–574. https://doi.org/10.1145/2702123.2702314

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*(2), 468–477. https://doi.org/10.3758/s13428-011-0064-1

Dautenhahn, K. (1998a). The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied Artificial Intelligence*, *12*(7-8), 573–617. https://doi.org/10.1080/088395198117550

Dautenhahn, K. (1998b). Embodiment and interaction in socially intelligent life-like agents. *Proceedings of the International Workshop on Computation for Metaphors, Analogy, and Agents*, 102–141. https://doi.org/10.1007/3-540-48834-0_7

Dautenhahn, K., Woods, S., Kaouri, C., Walters, M. L., Koay, K. L., & Werry, I. (2005). What is a robot companion - friend, assistant or butler? *Proceedings of the International Conference on Intelligent Robots and Systems*, 1192–1197. https://doi.org/10.1109/IROS.2005.1545189

Davis, M., & Hadiks, D. (1994). Nonverbal aspects of therapist attunement. *Journal of Clinical Psychology*, *50*(3), 393–405. https://doi.org/10.1002/1097-4679(199405)50:3<393::AID-JCLP2270500311>3.0.CO;2-T

De Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience*, *6*(6), 463–475. https://doi.org/10.1038/nrn1683

Deaton, J. E., Barba, C., Santarelli, T., Rosenzweig, L., Souders, V., McCollum, C., Seip, J., Knerr, B. W., & Singer, M. J. (2005). Virtual environment cultural training for operational readiness (VECTOR). *Virtual Reality*, *8*(3), 156–167. https://doi.org/10.1007/s10055-004-0145-x

Deladisma, A. M., Cohen, M., Stevens, A., Wagner, P., Lok, B., Bernard, T., Oxendine, C., Schumacher, L., Johnsen, K., Dickerson, R., Raij, A., Wells, R., Duerson, M., Harper, J. G., & Lind, D. S. (2007). Do medical students respond empathetically to a virtual patient? *The American Journal of Surgery*, *193*(6), 756–760. https://doi.org/10.1016/j.amjsurg.2007.01.021

DeLongis, A., Folkman, S., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology, 54*(3), 486–495. https://doi.org/10.1037/0022-3514.54.3.486

de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, *106*(1), 73–88. https://doi.org/10.1037/a0034251

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., & Morency, L.-P. (2014). Simsensei kiosk: A virtual human interviewer for healthcare decision support. *Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems*, 1061–1068.

Dias, J., Mascarenhas, S., & Paiva, A. (2014). FAtiMA modular: Towards an agent architecture with a generic appraisal framework. In T. Bosse, J. Broekens, J. Dias, & J. van der Zwaan (Eds.), *Emotion modeling* (pp. 44–56). https://doi.org/10.1007/978-3-319-12973-0_3

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*(4), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014). The 'hedonic' in human-computer interaction: History, contributions, and future research directions. *Proceedings of the 2014 Conference on Designing Interactive Systems*, 305–314. https://doi.org/10.1145/2598510.2598549

Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. *Research in Personnel and Human Resources Management*, *12*(1), 79–123.

D'Mello, S., & Kory, J. (2012). Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, 31–38. https://doi.org/10.1145/2388676.2388686

Donnelly, D. A., & Murray, E. J. (1991). Cognitive and emotional changes in written essays and therapy interviews. *Journal of Social and Clinical psychology*, *10*(3), 334–350. https://doi.org/10.1521/jscp.1991.10.3.334

Eckberg, D. L. (1983). Human sinus arrhythmia as an index of vagal cardiac outflow. *Journal of Applied Physiology*, *54*(4), 961–966. https://doi.org/10.1152/jappl.1983.54.4.961

Einstein, D., & Lanning, K. (1998). Shame, guilt, ego development and the five-factor model of personality. *Journal of Personality*, *66*(4), 555–582. https://doi.org/10.1111/1467-6494.00024

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98. https://doi.org/10.1515/semi.1969.1.1.49

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Elison, J., Lennon, R., & Pulos, S. (2006). Investigating the compass of shame: The development of the compass of shame scale. *Social Behavior and Personality: An International Journal*, *34*(3), 221–238. https://doi.org/10.2224/sbp.2006.34.3.221

Endrass, B., André, E., Rehm, M., & Nakano, Y. (2013). Investigating culture-related aspects of behavior for virtual characters. *Autonomous Agents and Multi-Agent Systems*, *27*(2), 277–304. https://doi.org/10.1007/s10458-012-9218-5

Exline, R., Gray, D., & Schuette, D. (1965). Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology*, *1*(3), 201–209. https://doi.org/10.1037/h0021865

Feldman Barrett, L. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of Personality and Social Psychology*, *87*(2), 266–281. https://doi.org/10.1037/0022-3514.87.2.266

Feldman Barrett, L. (2017). *How emotions are made: The secret life of the brain.* Houghton Mifflin Harcourt.

Fessler, D. M. T. (2007). From appeasement to conformity: Evolutionary and cultural perspectives on shame competition, and cooperation. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 174–193). Guilford Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

Fischer, A. H., & Manstead, A. S. (2008). Social functions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 456–468). Guilford Press.

Fonagy, P., Gergely, G., & Jurist, E. L. (2018). *Affect regulation, mentalization and the development of the self.* Routledge.

Forsythe, S. M. (1990). Effect of applicant's clothing on interviewer's decision to hire. *Journal of Applied Social Psychology*, *20*(19), 1579–1595. https://doi.org/10.1111/j.1559-1816.1990.tb01494.x

Foster, M. E., Avramides, K., Bernardini, S., Chen, J., Frauenberger, C., Lemon, O., & Porayska-Pomsta, K. (2010). Supporting children's social communication skills through interactive narratives with virtual characters. *Proceedings of the 18th ACM International Conference on Multimedia*, 1111–1114. https://doi.org/10.1145/1873951.1874163

Frates, E. P., Moore, M. A., Lopez, C. N., & McMahon, G. T. (2011). Coaching for behavior change in physiatry. *American Journal of Physical Medicine & Rehabilitation*, *90*(12), 1074–1082. https://doi.org/10.1097/PHM.0b013e31822dea9a

Freeman, G. L., Manson, G., Katzoff, E., & Pathman, J. (1942). The stress interview. *The Journal of Abnormal and Social Psychology*, *37*(4), 427–447. https://doi.org/10.1037/h0059025

Frijda, N. H. (1987). *The emotions*. Cambridge University Press.

Frijda, N. H. (2008). The psychologists' point of view. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 456–468). Guilford Press.

Fu, F., Tarnita, C. E., Christakis, N. A., Wang, L., Rand, D. G., & Nowak, M. A. (2012). Evolution of in-group favoritism. *Scientific Reports*, *2*(1), 1–6. https://doi.org/10.1038/srep00460

Fylan, F. (2005). Semi-structured interviewing. In J. Miles & P. Gilbert (Eds.), *A handbook of research methods for clinical and health psychology* (pp. 65–78). Oxford University Press. https://doi.org/10.1093/med:psych/9780198527565.003.0006

Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. NYU press.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*(1), 71–85. https://doi.org/10.30658/hmc.1.5

Gebhard, P. (2005). ALMA – a layered model of affect. *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, 29–36. https://doi.org/10.1145/1082473.1082478

Gebhard, P., Baur, T., Damian, I., Mehlmann, G., Wagner, J., & André, E. (2014). Exploring interaction strategies for virtual characters to induce stress in simulated job interviews. *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*, 661–668.

Gebhard, P., Mehlmann, G., & Kipp, M. (2012). Visual scenemaker–a tool for authoring interactive virtual characters. *Journal on Multimodal User Interfaces*, *6*, 3–11. https://doi.org/10.1007/s12193-011-0077-1

Gebhard, P., Schneeberger, T., André, E., Baur, T., Damian, I., Mehlmann, G., König, C., & Langer, M. (2019a). Serious games for training social skills in job interviews. *IEEE Transactions on Games*, *11*(4), 340–351. https://doi.org/10.1109/TG.2018.2808525

Gebhard, P., Schneeberger, T., Baur, T., & André, E. (2018). MARSSI: Model of appraisal, regulation, and social signal interpretation. *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, 497–506.

Gebhard, P., Schneeberger, T., Dietz, M., André, E., & Bajwa, N. u. H. (2019b). Designing a mobile social and vocational reintegration assistant for burn-out outpatient treatment. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 13–15. https://doi.org/10.1145/3308532.3329460

Gebhard, P., Schneeberger, T., Mehlmann, G., Baur, T., & André, E. (2019c). Designing the impression of social agents' real-time interruption handling. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 19–21. https://doi.org/10.1145/3308532.3329435

Geiskkovitch, D. Y., Cormier, D., Seo, S. H., & Young, J. E. (2016). Please continue, we need more data: An exploration of obedience to robots. *Journal of Human-Robot Interaction*, *5*(1), 82–99. https://doi.org/10.5898/JHRI.5.1

Geisler, F. C., Kubiak, T., Siewert, K., & Weber, H. (2013). Cardiac vagal tone is associated with social engagement and self-regulation. *Biological Psychology*, *93*(2), 279–286. https://doi.org/10.1016/j.biopsycho.2013.02.013

Goessl, V. C., Curtiss, J. E., & Hofmann, S. G. (2017). The effect of heart rate variability biofeedback training on stress and anxiety: A meta-analysis. *Psychological Medicine*, *47*(15), 2578–2586. https://doi.org/10.1017/S0033291717001003

Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots*, *39*(3), 293–312. https://doi.org/10.1007/s10514-015-9457-9

Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, *9*(2), 205–217. https://doi.org/10.1007/BF00148127

Gratch, J., Kang, S.-H., & Wang, N. (2013). Using social agents to explore theories of rapport and emotional resonance. In J. Gratch & S. Marsella (Eds.), *Social emotions in nature and artifact* (pp. 181–198). Oxford University Press.

Gratch, J., & Marsella, S. (2005). Evaluating a computational model of emotion. *Autonomous Agents and Multi-Agent Systems*, *11*(1), 23–43. https://doi.org/10.1007/s10458-005-1081-1

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., & Morency, L.-P. (2006). Virtual rapport. *Proceedings of the 6th International Conference on Intelligent Virtual Agents*, 14–27. https://doi.org/10.1007/11821830_2

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. *Proceedings of the 7th International Conference on Intelligent Virtual Agents*, 125–138. https://doi.org/10.1007/978-3-540-74997-4_12

Gross, J. J. (2013). Emotion regulation: Conceptual and empirical foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–22). Guilford Publications.

Hahn, W. K. (2001). The experience of shame in psychotherapy supervision. *Psychotherapy: Theory, Research, Practice, Training*, *38*(3), 272–282. https://doi.org/10.1037/0033-3204.38.3.272

Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social Development*, *10*(1), 79–119. https://doi.org/10.1111/1467-9507.00150

Hall, L., Jones, S., Paiva, A., & Aylett, R. (2009). FearNot! Providing children with strategies to cope with bullying. *Proceedings of the 8th International Conference on Interaction Design and Children*, 276–277. https://doi.org/10.1145/1551788.1551854

Hall, L., Tazzyman, S., Hume, C., Endrass, B., Lim, M.-Y., Hofstede, G., Paiva, A., Andre, E., Kappas, A., & Aylett, R. (2015). Learning to overcome cultural conflict through engaging with intelligent agents in synthetic cultures. *International Journal of Artificial Intelligence in Education*, *25*(2), 291–317. https://doi.org/10.1007/s40593-014-0031-y

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology*, *1*(1), 293–319. https://doi.org/10.1146/annurev.clinpsy.1.102803.143938

Hansen, A. L., Johnsen, B. H., Sollers, J. J., Stenvik, K., & Thayer, J. F. (2004). Heart rate variability and its relation to prefrontal cognitive function: The effects of training and detraining. *European Journal of Applied Physiology*, *93*(3), 263–272. https://doi.org/10.1007/s00421-004-1208-0

Hansen, L. (2013). *8 drivers who blindly followed their GPS into disaster.* https://theweek.com/articles/464674/8-drivers-who-blindly-followed–gps-into-disaster.

Harder, D. W., & Greenwald, D. F. (1999). Further validation of the shame and guilt scales of the harder personal feelings questionnaire-2. *Psychological Reports*, *85*(1), 271–281. https://doi.org/10.2466/pr0.1999.85.1.271

Harris, P. L. (1985). What children know about the situations that provoke emotion. In M. Lewis & C. Saarni (Eds.), *The socialization of emotions* (pp. 161–185). Springer.

Hartholt, A., Mozgai, S., Fast, E., Liewer, M., Reilly, A., Whitcup, W. R., & Rizzo, A. S. (2019). Virtual humans in augmented reality: A first step towards real-world embedded virtual roleplayers. *Proceedings of the 7th International Conference on Human-Agent Interaction*, 205–207. https://doi.org/10.1145/3349537.3352766

Heimberg, R. G., Keller, K. E., & Peca-Baker, T. A. (1986). Cognitive assessment of social-evaluative anxiety in the job interview: Job interview self-statement schedule. *Journal of Counseling Psychology*, *33*(2), 190–195.

Heinz, B. (2003). Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics*, *35*(7), 1113–1142. https://doi.org/10.1016/S0378-2166(02)00190-X

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review*, *17*(2), 142–157. https://doi.org/10.1177/1088868312472607

Hoffmann, L., Krämer, N. C., Lam-Chi, A., & Kopp, S. (2009). Media equation revisited: Do users show polite reactions towards an embodied agent? *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, 159–165. https://doi.org/http://dx.doi.org/10.1007/978-3-642-04380-2_19

Hopkins, I. M., Gower, M. W., Perez, T. A., Smith, D. S., Amthor, F. R., Casey Wimsatt, F., & Biasini, F. J. (2011). Avatar assistant: Improving social skills in students with an asd through a computer-based intervention. *Journal of Aautism and Developmental Disorders*, *41*(11), 1543–1555. https://doi.org/10.1007/s10803-011-1179-z

Hoque, M. E. (2012). My automated conversation helper (MACH): Helping people improve social skills. *Proceedings of the International Conference on Multimodal Interaction*, 313–316. https://doi.org/10.1145/2388676.2388745

Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., & Picard, R. W. (2013). MACH: My automated conversation coach. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 697–706. https://doi.org/10.1145/2493432.2493502

Hwang, H., & Matsumoto, D. (2019). Functions of emotions. In R. Biswas-Diener & E. Diener (Eds.), *Noba textbook series: Psychology*. DEF publishers. http://noba.to/w64szjxu

Imada, A. S., & Hakel, M. D. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology*, *62*(3), 295–300. https://doi.org/10.1037/0021-9010.62.3.295

Inkpen, K. M., & Sedlins, M. (2011). Me and my avatar: Exploring users' comfort with avatars for workplace communication. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 383–386. https://doi.org/10.1145/1958824.1958883

Islam, T., Mittal, U., Nimal, A., & Sharma, M. (2014). High frequency surface acoustic wave (SAW) device for toxic vapor detection: Prospects and challenges. In A. Mason, S. Mukhopadhyay, K. Jayasundera, & N. Bhattacharyya (Eds.), *Sensing technology: Current status and future trends II* (pp. 217–241). https://doi.org/10.1007/978-3-319-02315-1_11

Izard, C. E. (1977). *Human emotions. Emotions, personality, and psychotherapy.* Plenum.

Izard, C. E., Dougherty, F., Bloxom, B., & Kotsch, N. (1993). *The differential emotions scale: A method of measuring the meaning of subjective experience of discrete emotions*. Vanderbilt University.

Jackson, S. E., Hall, N. C., Rowe, P. M., & Daniels, L. M. (2009). Getting the job: Attributional retraining and the employment interview. *Journal of Applied Social Psychology*, *39*(4), 973–998. https://doi.org/10.1111/j.1559-1816.2009.00468.x

Jansen, A., König, C. J., Stadelmann, E. H., & Kleinmann, M. (2012). Applicants' self-presentational behavior: What do recruiters expect and what do they get? *Journal of Personnel Psychology*, *11*(2), 77–85. https://doi.org/10.1027/1866-5888/a000046

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory – versions 4a and 5.* University of California, Institute of Personality; Social Research.

Johnsen, K., Dickerson, R., Raij, A., Harrison, C., Lok, B., Stevens, A., & Lind, D. S. (2006). Evolving an immersive medical communication skills trainer. *Presence*, *15*(1), 33–46. https://doi.org/10.1162/pres.2006.15.1.33

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated Pedagogical Agents: Face-to-face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, *11*(1), 47–78.

Johnson, W. L., & Valente, A. (2008). Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 1632–1639. https://doi.org/10.1609/aimag.v30i2.2240

Kächele, M., Rukavina, S., Palm, G., Schwenker, F., & Schels, M. (2015). Paradigms for the construction and annotation of emotional corpora for real-world human-computer-interaction. *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, 367–373. https://doi.org/10.5220/0005282703670373

Kaiser, S., & Wehrle, T. (2001). The role of facial expression in intra-individual and inter-individual emotion regulation. *AAAI Technical Report*. https://doi.org/https://www.aaai.org/Papers/Symposia/Fall/2001/FS-01-02/FS01-02-014.pdf

Kang, S.-H., & Gratch, J. (2010). Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds*, *21*(3-4), 473–482. https://doi.org/10.1002/cav.345

Kapoor, A., & Picard, R. W. (2005). Multimodal affect recognition in learning environments. *Proceedings of the 13th Annual ACM International Conference on Multimedia*, 677–682. https://doi.org/10.1145/1101149.1101300

Karthikeyan, P., Murugappan, M., & Yaacob, S. (2013). Detection of human stress using short-term ECG and HRV signals. *Journal of Mechanics in Medicine and Biology*, *13*(2), Article 1350038. https://doi.org/10.1142/S0219519413500383

Kavakli, M., Li, M., & Rudra, T. (2012). Towards the development of a virtual counselor to tackle students' exam stress. *Journal of Integrated Design and Process Science*, *16*(1), 5–26. https://doi.org/10.3233/jid-2012-0004

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems*, *2*(1), 26–41. https://doi.org/10.1145/357417.357420

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, *68*(3), 441–454.

Keltner, D. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, *10*(2), 155–172. https://doi.org/10.1080/026999396380312

Keltner, D. (2003). Expression and the course of life: Studies of emotion, personality, and psychopathology from a social-functional perspective. *Annals of the New York Academy of Sciences*, *1000*(1), 222–243. https://doi.org/10.1196/annals.1280.011

Kenny, P. G., Parsons, T. D., Gratch, J., & Rizzo, A. A. (2008). Evaluation of Justina: A virtual patient with PTSD. *Proceedings of the 8th ACM International Conference on Intelligent Virtual Agents*, 394–408. https://doi.org/10.1007/978-3-540-85483-8_40

Kerr, S., Neale, H., & Cobb, S. (2002). Virtual environments for social skills training: The importance of scaffolding in practice. *Proceedings of the ACM Conference on Assistive Technologies*, 104–110. https://doi.org/10.1145/638249.638269

Kiesler, S., & Sproull, L. (1997). "Social" human-computer interaction. In B. Friedman (Ed.), *Human Values and the Design of Computer Technology* (pp. 191–199). Cambridge University Press.

Kim, J. M., Hill Jr, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., Marsella, S., Pynadath, D., & Hart, J. (2009). Bilat: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, *19*(3), 289–308.

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier social stress test' – a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1-2), 76–81. https://doi.org/10.1159/000119004

Kivimäki, M., Virtanen, M., Elovainio, M., Kouvonen, A., Väänänen, A., & Vahtera, J. (2006). Work stress in the etiology of coronary heart disease–a meta-analysis. *Scandinavian Journal of Work, Environment & Health*, *32*(6), 431–442.

Klein, E. M., Brähler, E., Dreier, M., Reinecke, L., Müller, K. W., Schmutzer, G., Wölfling, K., & Beutel, M. E. (2016). The German version of the perceived stress scale–psychometric characteristics in a representative German community sample. *BMC Psychiatry*, *16*(1), Article 159. https://doi.org/10.1186/s12888-016-0875-9

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide: Design and evaluation of a real-world application. *Proceedings of the 5th International Conference on Intelligent Virtual Agents*, 329–343. https://doi.org/10.1007/11550617_28

Kory-Westlund, J. M., Won Park, H., Grover, I., & Breazeal, C. (2022). Long-term interaction with relational SIAs. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 2: Interactivity, platforms, application* (pp. 195–260). Association for Computing Machinery. https://doi.org/10.1145/3563659.3563667

Kraag, G., Zeegers, M. P., Kok, G., Hosman, C., & Abu-Saad, H. H. (2006). School programs targeting stress management in children and adolescents: A meta-analysis. *Journal of School Psychology*, *44*(6), 449–472. https://doi.org/10.1016/j.jsp.2006.07.001

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*, 569–598. https://doi.org/10.1146/annurev.psych.50.1.569

Krämer, N. C., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you: The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies*, *71*(3), 335–349. https://doi.org/10.1016/j.ijhcs.2012.09.006

Krämer, N. C., Lucas, G., Schmitt, L., & Gratch, J. (2018). Social snacking with a virtual agent – on the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human-Computer Studies*, *109*(1), 112–121. https://doi.org/10.1016/j.ijhcs.2017.09.001

Krämer, N. C., & Manzeschke, A. (2021). Social reactions to socially interactive agents and their ethical implications. In B. Lugrin, C. Pelachaud, & D.

Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 1: Methods, behavior, cognition* (pp. 77–104). Association for Computing Machinery. https://doi.org/10.1145/3477322.3477326

Krantz, D. S., & McCeney, M. K. (2002). Effects of psychological and social factors on organic disease: A critical assessment of research on coronary heart disease. *Annual Review of Psychology*, *53*(1), 341–369. https://doi.org/10.1146/annurev.psych.53.100901.135208

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, *37*(9), 1539–1553. https://doi.org/10.1037/0022-3514.37.9.1539

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS). *Diagnostica*, *42*(2), 139–156.

Kromand, D. (2007). Avatar categorization. *Proceedings of Digital Games Research Association Conference*, 400–406.

Krupp, J., Taubner, S., Huber, D., & Hamburger, A. (2019). Validierung der deutschen Uebersetzung der Psychological Mindedness Scale (PMS). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, *65*(1), 27–41. https://doi.org/10.13109/zptm.2019.65.1.27

Kudielka, B. M., Hellhammer, D. H., & Kirschbaum, C. (2007). Ten years of research with the Trier social stress test–revisited. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 56–83). The Guilford Press.

Lagos, L., Vaschillo, E., Vaschillo, B., Lehrer, P. M., Bates, M., & Pandina, R. (2011). Virtual reality-assisted heart rate variability biofeedback as a strategy to improve golf performance: A case study. *Biofeedback*, *39*(1), 15–20. https://doi.org/10.5298/1081-5937-39.1.11

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162. https://doi.org/10.1023/A:1025389814290

Lane, H. C., Hays, M. J., Core, M. G., & Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology*, *105*(4), 1026–1035. https://doi.org/10.1037/a0031506

Lane, H. C., Hays, M., Core, M., Gomboc, D., Forbell, E., Auerbach, D., & Rosenberg, M. (2008). Coaching intercultural communication in a serious game. *Proceedings of the 16th International Conference on Computers in Education*, 35–42.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International affective picture system (IAPS): Technical manual and affective ratings*. NIMH Center for the Study of Emotion & Attention.

Langer, M., & König, C. J. (2018). Introducing and testing the creepiness of situation scale (CRoSS). *Frontiers in Psychology*, *9*, Article 2220. https://doi.org/10.3389/fpsyg.2018.02220

Langer, M., König, C. J., Gebhard, P., & André, E. (2016). Dear computer, teach me manners: Testing virtual employment interview training. *International Journal of Selection and Assessment*, *24*(4), 312–323. https://doi.org/10.1111/ijsa.12150

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Symposium of the Austrian HCI and Usability Engineering Group*, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

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

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology*, *44*(1), 1–22. https://doi.org/10.1146/annurev.ps.44.020193.000245

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer.

Lehrer, P. M. (2007). Biofeedback training to increase heart rate variability. In P. M. Lehrer, R. L. Woolfolk, & W. E. Sime (Eds.), *Principles and practice of stress management* (3rd ed., pp. 227–248). The Guilford Press.

Lehrer, P. M. (2013). How does heart rate variability biofeedback work? Resonance, the baroreflex, and other mechanisms. *Biofeedback*, *41*(1), 26–31. https://doi.org/10.5298/1081-5937-41.1.02

Lemaire, J. B., Wallace, J. E., Lewin, A. M., de Grood, J., & Schaefer, J. P. (2011). The effect of a biofeedback-based stress management tool on physician stress: A randomized controlled clinical trial. *Open Medicine*, *5*(4), 154–165.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems*, 359–366.

Leudar, I., Costall, A., & Francis, D. (2004). Theory of mind: A critical assessment. *Theory & Psychology*, *14*(5), 571–578. https://doi.org/10.1177/0959354304046173

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241–293. https://doi.org/10.1111/peps.12052

Lewis, M. (1992). *Shame: The exposed self*. The Free Press.

Lewis, M. (2008). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 742–756). The Guilford Press.

Lisetti, C. L., & Nasoz, F. (2002). MAUI: A multimodal affective user interface. *Proceedings of the 10th ACM International Conference on Multimedia*, 161–170. https://doi.org/10.1145/641007.641038

Lok, B., Ferdig, R. E., Raij, A., Johnsen, K., Dickerson, R., Coutts, J., Stevens, A., & Lind, D. S. (2006). Applying virtual reality in medical communication education: Current findings and potential teaching and learning benefits of immersive virtual patients. *Virtual Reality*, *10*(3), 185–195. https://doi.org/10.1007/s10055-006-0037-3

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. https://doi.org/10.1016/j.chb.2014.04.043

Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L.-P. (2017). Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, *4*, Article 51. https://doi.org/10.3389/frobt.2017.00051

Lugrin, B. (2021). Introduction to socially interactive agents. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 1: Methods, behavior, cognition* (pp. 77–104). Association for Computing Machinery. https://doi.org/10.1145/3477322.3477326

Lugrin, B., & Rehm, M. (2021). Culture for socially interactive agents. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 1: Methods, behavior, cognition* (pp. 77–104). Association for Computing Machinery. https://doi.org/10.1145/3477322.3477326

Lunney, A., Cunningham, N. R., & Eastin, M. S. (2016). Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Computers in Human Behavior*, *65*(1), 114–120. https://doi.org/10.1016/j.chb.2016.08.007

Malik, M. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Task force of the European society of cardiology and the North American society for pacing and electrophysiology. *Circulation*, *93*(5), 1043–1065.

Marsella, S., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, *10*(1), 70–90. https://doi.org/10.1016/j.cogsys.2008.03.005

Marsella, S., & Gratch, J. (2014). Computationally Modeling Human Emotion. *Communications of the ACM*, *57*(12), 56–67. https://doi.org/10.1145/2631912

Marsella, S., Gratch, J., & Petta, P. (2010). Computational Models of Emotion. In K. R. Scherer, T. Bänzinger, & E. B. Roesch (Eds.), *Blueprint for affective computing (A sourcebook)* (pp. 21–41). Oxford University Press.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger state–trait anxiety inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Masaoka, Y., & Homma, I. (1997). Anxiety and respiratory patterns: Their relationship during mental stress and physical load. *International Journal of Psychophysiology*, *27*(2), 153–159. https://doi.org/10.1016/S0167-8760(97)00052-4

Mascarenhas, S., Silva, A., Paiva, A., Aylett, R., Kistler, F., André, E., Degens, N., Hofstede, G. J., & Kappas, A. (2013). Traveller: An intercultural training system with intelligent agents. *Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems*, 1387–1388.

Maurer, R. E., & Tindall, J. H. (1983). Effect of postural congruence on client's perception of counselor empathy. *Journal of Counseling Psychology*, *30*(2), 158–163. https://doi.org/10.1037/0022-0167.30.2.158

Mazure, C. M. (1998). Life stressors as risk factors in depression. *Clinical Psychology: Science and Practice*, *5*(3), 291–313. https://doi.org/10.1111/j.1468-2850.1998.tb00151.x

McCarthy, J., & Goffin, R. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology*, *57*(3), 607–637. https://doi.org/10.1111/j.1744-6570.2004.00002.x

McNaughton, D., Hamlin, D., McCarthy, J., Head-Reeves, D., & Schreiner, M. (2008). Learning to listen: Teaching an active listening strategy to preservice education professionals. *Topics in Early Childhood Special Education*, *27*(4), 223–231. https://doi.org/10.1177/0271121407311241

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261–292. https://doi.org/10.1007/BF02686918

Menne, I. M. (2017). Yes, of course? An investigation on obedience and feelings of shame towards a robot. *Proceedings of the 9th International Conference on Social Robotics*, 365–374. https://doi.org/10.1007/978-3-319-70022-9_36

Menne, I. M., & Lugrin, B. (2017). In the face of emotion: A behavioral study on emotions towards a robot using the facial action coding system. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 205–206. https://doi.org/10.1145/3029798.3038375

Merten, J. (1996). *Affekte und die Regulation nonverbalen, interaktiven Verhaltens: Strukturelle Aspekte des mimisch-affektiven Verhaltens und die Integration von Affekten in Regulationsmodelle*. Lang.

Merten, J. (2003). Context-analysis of facial-affective behavior in clinical populations. In M. Katsikitis (Ed.), *The human face. Measurment and meaning* (pp. 131–147). https://doi.org/10.1007/978-1-4615-1063-5_7

Merten, J., & Krause, R. (1993). *DAS (Differentielle Affekt Skala)*. Fachbereich Sozial und Umweltwissenschaften / Fachrichtung Psychologie.

Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review*, *6*(4), 298–302. https://doi.org/10.1177/1754073914534480

Mikosch, P., Hadrawa, T., Laubreiter, K., Brandl, J., Pilz, J., Stettner, H., & Grimm, G. (2010). Effectiveness of respiratory-sinus-arrhythmia biofeedback on state-anxiety in patients undergoing coronary angiography. *Journal of Advanced Nursing*, *66*(5), 1101–1110. https://doi.org/10.1111/j.1365-2648.2010.05277.x

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, *18*(1), 57–76. https://doi.org/10.1177/001872676\\501800105

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. Tavistock Publications.

Milne, M., Luerssen, M. H., Lewis, T. W., Leibbrandt, R. E., & Powers, D. M. W. (2010). Development of a virtual agent based social tutor for children with autism spectrum disorders. *Proceedings of the 2010 International Joint Conference on Neural Networks*, 1–9. https://doi.org/10.1109/IJCNN.2010.5596584

Monroe, S. M., & Simons, A. D. (1991). Diathesis–stress theories in the context of life stress research implications for the depressive disorders. *Psychological Bulletin*, *110*(3), 406–425. https://doi.org/10.1037/0033-2909.110.3.406

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mortillaro, M., Meuleman, B., & Scherer, K. R. (2012). Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions*, *3*(1), 18–32. https://doi.org/http://dx.doi.org/10.4018/jse.2012010102

Moser, U. (2009). *Theorie der Abwehrprozesse: Die mentale Organisation psychischer Störungen*. Brandes & Apsel.

Moser, U. (2013). *Von der Schwierigkeit, die Brust an den richtigen Ort zu setzen: Naive, implizite und explizite Reflexivität*. Brandes & Apsel Verlag.

Moser, U., & von Zeppelin, I. (1991). *Cognitive-affective processes: New ways of psychoanalytic modeling*. Springer.

Moser, U., & von Zeppelin, I. (1996). Die Entwicklung des Affektsystems. *Psyche - Zeitschrift für Psychoanalyse und ihre Anwendungen*, *50*(1), 32–84.

Moshkina, L., Trickett, S., & Trafton, J. G. (2014). Social engagement in public places: A tale of one robot. *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 382–389. https://doi.org/10.1145/2559636.2559678

Mower, E., Black, M. P., Flores, E., Williams, M., & Narayanan, S. (2011). Rachel: Design of an emotionally targeted interactive agent for children with autism. *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo*, 1–6. https://doi.org/10.1109/ICME.2011.6011990

Munafo, M., Patron, E., & Palomba, D. (2016). Improving managers' psychophysical well-being: Effectiveness of respiratory sinus arrhythmia biofeedback. *Applied Psychophysiology and Biofeedback*, *41*(2), 129–139. https://doi.org/10.1007/s10484-015-9320-y

Murphy, K. P. (2002). Dynamic Bayesian Networks. *Probabilistic Graphical Models*, *7*.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology*, *29*(5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C., Moon, Y., Morkes, J., Kim, E.-Y., & Fogg, B. J. (1997). Computers are social actors: A review of current research. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 137–162). Center for the Study of Language; Information.

Nass, C., & Steuer, J. (2006). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research*, *19*(4), 504–527. https://doi.org/10.1111/j.1468-2958.1993.tb00311.x

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78.

Nathanson, D. L. (1994). *Shame and Pride: Affect, Sex, and the Birth of the Self*. WW Norton & Company.

Nazir, A., Krumhuber, E., Degens, D., Endrass, B., Hume, C., Swiderska, A., Hodgson, J., Ritter, C., Mascarenhas, S., & Aylett, R. (2012). ECUTE:

Difference is good. *Proceedings of the International Conference E-Learning*, 425–429.

Neumayr, M. (2018). *Der Einfluss von Mimikry auf virtuelle Einstellungsinterviews* (Master's thesis). Saarland University.

Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C. (2009). Greta: An interactive expressive ECA system. *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, 1399–1400.

Noh, S. R., & Isaacowitz, D. M. (2013). Emotional faces in context: Age differences in recognition accuracy and scanning patterns. *Emotion*, *13*(2), 238–246. https://doi.org/10.1037/a0030234

Nordstrom, C. R., Williams, K. B., & LeBreton, J. M. (1996). The effect of cognitive load on the processing of employment selection information. *Basic and Applied Social Psychology*, *18*(3), 305–318. https://doi.org/10.1207/s15324834basp1803_4

Nussinovitch, U., Elishkevitz, K. P., Katz, K., Nussinovitch, M., Segev, S., Volovitz, B., & Nussinovitch, N. (2011). Reliability of ultra-short ECG indices for heart rate variability. *Annals of Noninvasive Electrocardiology*, *16*(2), 117–122. https://doi.org/10.1111/j.1542-474X.2011.00417.x

Ochs, M., Mestre, D., De Montcheuil, G., Pergandi, J.-M., Saubesty, J., Lombardo, E., Francon, D., & Blache, P. (2019). Training doctors' social skills to break bad news: Evaluation of the impact of virtual environment displays on the sense of presence. *Journal on Multimodal User Interfaces*, *13*(1), 41–51. https://doi.org/10.1007/s12193-018-0289-8

Ojha, S., Vitale, J., & Williams, M.-A. (2021). Computational emotion models: A thematic review. *International Journal of Social Robotics*, *13*(6), 1253–1279. https://doi.org/10.1007/s12369-020-00713-1

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Parkinson, B., & Manstead, A. S. R. (2015). Current emotion research in social psychology: Thinking about emotions and other people. *Emotion Review*, *7*(4), 371–380. https://doi.org/10.1177/1754073915590624

Paul, M., & Garg, K. (2012). The effect of heart rate variability biofeedback on performance psychology of basketball players. *Applied Psychophysiology and Biofeedback*, *37*(2), 131–144. https://doi.org/10.1007/s10484-012-9185-2

Pauw, L. S., Sauter, D. A., van Kleef, G. A., Lucas, G. M., Gratch, J., & Fischer, A. H. (2022). The avatar will see you now: Support from a virtual human provides socio-emotional benefits. *Computers in Human Behavior*, *136*, Article 107368. https://doi.org/10.1016/j.chb.2022.107368

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review*, *26*(3), 894–900. https://doi.org/10.3758/s13423-019-01571-3

Pelachaud, C. (2009). Studies on gesture expressivity for a virtual agent. *Speech Communication*, *51*(7), 630–639. https://doi.org/10.1016/j.specom.2008.04.009

Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., & Poggi, I. (2005). A model of attention and interest using gaze behavior. *Proceedings of the 5th International Conference on Intelligent Virtual Agents*, 229–240. https://doi.org/10.1007/11550617_20

Pfeifer, R. (1988). Artificial Intelligence Models of Emotion. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 287–320). Springer Dordrecht. https://doi.org/10.1007/978-94-009-2792-6

Picard, R. W. (1997). *Affective computing*. MIT Press.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *Transactions on Pattern Analysis & Machine Intelligence*, *23*(10), 1175–1191. https://doi.org/10.1109/34.954607

Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior*, *65*(1), 23–30. https://doi.org/10.1016/j.chb.2016.08.004

Polit, D. F., & Beck, C. T. (2009). *Essentials of nursing research: Appraising evidence for nursing practice*. Lippincott Williams & Wilkins.

Prada, R., & Rato, D. (2022). Socially interactive agents in games. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics. Volume 2: Interactivity, platforms, application* (pp. 493–526). Association for Computing Machinery. https://doi.org/10.1145/3563659.3563675

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/http://dx.doi.org/10.1017/S0140525X00076512

Prinsloo, G. E., Derman, W. E., Lambert, M. I., & Rauch, H. (2013). The effect of a single episode of short duration heart rate variability biofeedback on measures of anxiety and relaxation states. *International Journal of Stress Management*, *20*(4), 391–411. https://doi.org/10.1037/a0034777

Prinsloo, G. E., Rauch, H. L., Lambert, M. I., Muench, F., Noakes, T. D., & Derman, W. E. (2011). The effect of short duration heart ratevariability (HRV) biofeedback on cognitive performance during laboratory induced cognitive stress. *Applied Cognitive Psychology*, *25*(5), 792–801. https://doi.org/10.1002/acp.1750

Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. *Proceedings of the 1st International Conference on Multi-Agent Systems*, 312–319.

Ratanasiripong, P., Kaewboonchoo, O., Ratanasiripong, N., Hanklang, S., & Chumchai, P. (2015). Biofeedback intervention for stress, anxiety, and depression among graduate students in public health nursing. *Nursing Research and Practice*, *2015*, Article 160746. https://doi.org/10.1155/2015/160746

Raybourn, E. M., Deagle, E., Mendini, K., & Heneghan, J. (2005). Adaptive thinking & leadership simulation game training for special forces officers. *Proceedings of the Interservice, Industry Training, Simulation & Education Conference*, 1–9.

Razavi, S. Z., Ali, M. R., Smith, T. H., Schubert, L. K., & Hoque, M. E. (2016). The LISSA virtual human and ASD teens: An overview of initial experiments. *Proceedings of the 16th International Conference on Intelligent Virtual Agents*, 460–463. https://doi.org/10.1007/978-3-319-47665-0_55

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

Renaud, P., & Blondin, J.-P. (1997). The stress of Stroop performance: Physiological and emotional responses to color–word interference, task pacing, and pacing speed. *International Journal of Psychophysiology*, *27*(2), 87–97. https://doi.org/10.1016/S0167-8760(97)00049-4

Retzinger, S. M. (1995). Identifying shame and anger in discourse. *American Behavioral Scientist*, *38*(8), 1104–1113.

Richardson, K. M., & Rothstein, H. R. (2008). Effects of occupational stress management intervention programs: A meta-analysis. *Journal of Occupational Health Psychology*, *13*(1), 69–93. https://doi.org/10.1037/1076-8998.13.1.69

Rickel, J. (2001). Intelligent virtual agents for education and training: Opportunities and challenges. *Proceedings of the 3rd International Workshop on Intelligent Virtual Agents*, 15–22. https://doi.org/10.1007/3-540-44812-8_2

Rockstroh, C., Blum, J., & Göritz, A. S. (2019). Virtual reality in the application of heart rate variability biofeedback. *International Journal of Human-Computer Studies*, *130*(1), 209–220. https://doi.org/10.1016/j.ijhcs.2019.06.011

Rodrigues, S. H., Mascarenhas, S. F., Dias, J., & Paiva, A. (2009). “I can feel it too!”: Emergent Empathic Reactions Between Synthetic Characters. *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 1–7.

Rodrıéguez, L.-F., & Ramos, F. (2014). Development of computational models of emotions for autonomous agents: A review. *Cognitive Computation*, *6*(3), 351–375. https://doi.org/10.1007/s12559-013-9244-x

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(1), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rozanski, A., Blumenthal, J. A., & Kaplan, J. (1999). Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation*, *99*(16), 2192–2217. https://doi.org/10.1161/01.CIR.99.16.2192

Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and ethical considerations. *Proceedings of the Irish Conference on Artificial Intelligence and Cognitive Science*, 104–115.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, S., & Norvig, P. (2002). *Artificial intelligence: A modern approach*. Prentice Hall.

Sapouna, M., Wolke, D., Vannini, N., Watson, S., Woods, S., Schneider, W., Enz, S., Hall, L., Paiva, A., André, E., Dautenhahn, K., & Aylett, R. (2010). Virtual learning intervention to reduce bullying victimization in primary school: A controlled trial. *Journal of Child Psychology and Psychiatry*, *51*(1), 104–112. https://doi.org/10.1111/j.1469-7610.2009.02137.x

Satish, P., Muralikrishnan, K., Balasubramanian, K., & Shanmugapriya. (2015). Heart rate variability changes during Stroop color and word test among genders. *Indian Journal of Physiology and Pharmacology*, *59*(1), 9–15.

Scheff, T. J., & Retzinger, S. M. (2000). Shame as the master emotion of everyday life. *Journal of Mundane Behavior*, *1*(3), 303–324.

Schek, E. J., Mantovani, F., Realdon, O., Dias, J., Paiva, A., Schramm-Yavin, S., & Pat-Horenczyk, R. (2017). Positive technologies for promoting emotion regulation abilities in adolescents. *eHealth 360: International Summit on eHealth*, 169–174. https://doi.org/10.1007/978-3-319-49655-9_23

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Schmid Mast, M., Kleinlogel, E. P., Tur, B., & Bachmann, M. (2018). The future of interpersonal skills development: Immersive virtual reality training with virtual humans. *Human Resource Development Quarterly*, *29*(2), 125–141. https://doi.org/10.1002/hrdq.21307

Schneeberger, T., Gebhard, P., Baur, T., & André, E. (2019a). Parley: A transparent virtual social agent training interface. *Proceedings of the 24th In-*

ternational Conference on Intelligent User Interfaces: Companion, 35–36. https://doi.org/10.1145/3308557.3308674

Schneeberger, T., Sauerwein, N., Anglet, M. S., & Gebhard, P. (2020). Developing a social biofeedback training system for stress management training. *Companion Publication of the 2020 International Conference on Multimodal Interaction*, 472–476. https://doi.org/10.1145/3395035.3425222

Schneeberger, T., Scholtes, M., Hilpert, B., Langer, M., & Gebhard, P. (2019b). Can social agents elicit shame as humans do? *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction*, 164–170. https://doi.org/10.1109/ACII.2019.8925481

Schuller, B., Marchi, E., Baron-Cohen, S., Lassalle, A., O'Reilly, H., Pigat, D., Robinson, P., Davies, I., Baltrusaitis, T., & Mahmoud, M. (2015). Recent developments and results of ASC-inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions. *Proceedings of the 3rd International Workshop on Intelligent Digital Games for Empowerment and Inclusion as part of the 20th ACM International Conference on Intelligent User Interfaces*, 9–16.

Schwab, F. (2000). *Affektchoreographien. Eine evolutionspsychologische Analyse von Grundformen mimisch-affektiver Interaktionsmuster* (Doctoral dissertation). Saarland University.

Schwartz, M. S., & Andrasik, F. (2017). *Biofeedback: A practitioner's guide* (4th ed.). The Guilford Press.

Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation. *The Journal of Clinical Endocrinology & Metabolism*, *6*(2), 117–230. https://doi.org/10.1210/jcem-6-2-117

Selye, H. (2013). *Stress in health and disease*. Butterworth.

Shamekhi, A., Trinh, H., Bickmore, T. W., DeAngelis, T. R., Ellis, T., Houlihan, B. V., & Latham, N. K. (2016). A virtual self-care coach for individuals with spinal cord injury. *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 327–328. https://doi.org/10.1145/2982142.2982199

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061

Sherlin, L., Gevirtz, R., Wyckoff, S., & Muench, F. (2009). Effects of respiratory sinus arrhythmia biofeedback versus passive biofeedback control. *International Journal of Stress Management*, *16*(3), 233–248. https://doi.org/10.1037/a0016047

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS One*, *1*(1), Article e39. https://doi.org/10.1371/journal.pone.0000039

Slater, M., Pertaub, D., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications*, *19*(2), 6–9. https://doi.org/10.1109/38.749116

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838.

Smith, M. J., Ginger, E. J., Wright, K., Wright, M. A., Taylor, J. L., Humm, L. B., Olsen, D. E., Bell, M. D., & Fleming, M. F. (2014). Virtual reality job interview training in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(10), 2450–2463. https://doi.org/10.1007/s10803-014-2113-y

Soleymani, M., Pantic, M., & Pun, T. (2012). Multimodal emotion recognition in response to videos. *Transactions on Affective Computing*, *3*(2), 211–223. https://doi.org/10.1109/T-AFFC.2011.37

Sommer, R. (1959). Studies in personal space. *Sociometry*, *22*(3), 247–260.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, *11*(2), 97–124. https://doi.org/10.1207/s15327051hci1102_1

Staal, M. A. (2004). Stress, cognition, and human performance: A literature review and conceptual framework. *NASA Technical Reports*. https://ntrs.nasa.gov/citations/20060017835

Stansfield, S., Shawver, D., Sobel, A., Prasad, M., & Tapia, L. (2000). Design and implementation of a virtual reality system and its application to training medical first responders. *Presence: Teleoperators and Virtual Environments*, *9*(6), 524–556. https://doi.org/10.1162/105474600300040376

Stern, D. N. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. Routledge. https://doi.org/10.4324/9780429482137

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C., DiPietro, M., Allen, B., Ferdig, R., Foti, S., Jackson, J., Shin, M., Cendan, J., Watson, R., Duerson, M., Lok, B., Cohen, M., Wagner, P., & Lind, S. (2006). The use of virtual patients to teach medical students history taking and communication skills. *The American Journal of Surgery*, *191*(6), 806–811. https://doi.org/10.1016/j.amjsurg.2006.03.002

Stewart, G. L., Dustin, S. L., Barrick, M. R., & Darnold, T. C. (2008). Exploring the handshake in employment interviews. *Journal of Applied Psychology*, *93*(5), 1139–1146. https://doi.org/10.1037/0021-9010.93.5.1139

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/009365000027006001

Sutarto, A. P., Wahab, M. N. A., & Zin, N. M. (2010). Heart rate variability (HRV) biofeedback: A new training approach for operator's performance enhancement. *Journal of Industrial Engineering and Management*, *3*(1), 176–198. https://doi.org/http://dx.doi.org/10.3926/jiem..v3n1.p176-198

Swartout, W. R., Gratch, J., Hill Jr., R. W., Hovy, E., Marsella, S., Rickel, J., & Traum, D. (2006). Toward virtual humans. *AI Magazine*, *27*(2), Article 96. https://doi.org/10.1609/aimag.v27i2.1883

Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability. *Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering*, 1366–1369. https://doi.org/10.1007/978-3-540-89208-3_324

Tamir, M. (2011). The maturing field of emotion regulation. *Emotion Review*, *3*(1), 3–7. https://doi.org/10.1177/1754073910388685

Tangney, J. P. (1999). The self-conscious emotions: Shame, guilt, embarrassment and pride. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 541–568). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch26

Tao, J., & Tan, T. (2005). Affective computing: A review. *Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction*, 981–995.

Tartaro, A., & Cassell, J. (2008). Playing with virtual peers: Bootstrapping contingent discourse in children with autism. *Proceedings of the 8th International Conference for the Learning Sciences*, 382–389.

Thordarson, A., & Vilhjälmsson, H. H. (2019). SoCueVR: Virtual reality game for social cue detection training. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 46–48. https://doi.org/10.1145/3308532.3329440

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, *1*(4), 285–293. https://doi.org/10.1207/s15327965pli0104_1

Tielman, M. L., Neerincx, M. ., Bidarra, R., Kybartas, B., & Brinkman, W.-P. (2017). A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. *Journal of Medical Systems*, *41*(8), Article 125. https://doi.org/10.1007/s10916-017-0771-y

Titchener, E. B. (1912). The schema of introspection. *The American Journal of Psychology*, *23*(4), 485–508. https://doi.org/10.2307/1413058

Tomkins, S. S. (1963). *Affect imagery consciousness: Volume II: The negative affects*. Travistock Publications.

Tomkins, S. S. (1984). Affect theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 163–195). Psychology Press.

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 173–191). The Guilford Press.

Traum, D., Swartout, W., Gratch, J., & Marsella, S. (2008). A virtual human dialogue model for non-team interaction. In L. Dybkjær & W. Minker (Eds.), *Recent trends in discourse and dialogue* (pp. 45–67). Springer Netherlands.

Trepte, S., & Reinecke, L. (2010). Avatar creation and video game enjoyment: Effects of life-satisfaction, game competitiveness, and identification with the avatar. *Journal of Media Psychology: Theories, Methods, and Applications*, *22*(4), 171–184. https://doi.org/10.1027/1864-1105/a000022

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, *90*(2), 288–307. https://doi.org/10.1037/0022-3514.90.2.288

Ullrich, P. M., & Lutgendorf, S. K. (2002). Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine*, *24*(3), 244–250. https://doi.org/10.1207/S15324796ABM2403_10

Ursula, H., & Shlomo, H. (2015). The influence of context on emotion recognition in humans. *Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 1–6. https://doi.org/http://dx.doi.org/10.1109/FG.2015.7284842

Valstar, M., Baur, T., Cafaro, A., Ghitulescu, A., Potard, B., Wagner, J., André, E., Durieu, L., Aylett, M., Dermouche, S., et al. (2016a). Ask Alice: An artificial retrieval of information agent. *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 419–420.

Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Lalanne, D., Torres, M., Scherer, S., Stratou, G., Cowie, R., & Pantic, M. (2016b). Avec 2016: Depression, mood, and emotion recognition workshop and challenge. *Proceedings of the International Workshop on Audio/Visual Emotion Challenge*, 3–10.

Van Mulken, S., André, E., & Müller, J. (1998). The persona effect: How substantial is it? *Proceedings of the People and Computers XIII*, 53–66. https://doi.org/10.1007/978-1-4471-3605-7_4

Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., & Slater, M. (2006). Building expression into virtual characters. *Proceedings of the Eurographics 2006 - State of the Art Reports*, 21–61.

Vogt, T., & André, E. (2005). Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. *Proceedings of the IEEE*

International Conference on Multimedia and Expo, 474–477. https://doi.org/10.1109/ICME.2005.1521463

Vogt, T., André, E., & Bee, N. (2008). EmoVoice - A framework for online recognition of emotions from voice. *Proceedings of the 4th International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*, 188–199.

von Scheve, C. (2010). Die emotionale Struktur sozialer Interaktion: Emotionsexpression und soziale Ordnungsbildung/The emotional structure of social interaction: The expression of emotion and the emergence of social order. *Zeitschrift für Soziologie*, *39*(5), 346–362. https://doi.org/10.1515/zfsoz-2010-0501

Wagner, J., Baur, T., Schiller, D., Zhang, Y., Schuller, B., Valstar, M., & André, E. (2018a). Show me what you've learned: Applying cooperative machine learning for the semi-automated annotation of social signals. *IJCAI/ECAI 2018 Workshop on Explainable Artificial Intelligence*, 171–177.

Wagner, J., Baur, T., Zhang, Y., Valstar, M. F., Schuller, B., & André, E. (2018b). *Applying Cooperative Machine Learning to Speed Up the Annotation of Social Signals in Large Multi-modal Corpora*. https://doi.org/http://dx.doi.org/10.48550/ARXIV.1802.02565

Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., & André, E. (2013). The social signal interpretation (SSI) framework: Multimodal signal processing and recognition in real-time. *Proceedings of the 21st ACM International Conference on Multimedia*, 831–834. https://doi.org/10.1145/2502081.2502223

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2019). "Do you trust me?" Increasing user-trust by integrating virtual agents in explainable AI interaction design. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 7–9. https://doi.org/10.1145/3308532.3329441

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2021). "Let me explain!": Exploring the potential of virtual agents in explainable AI interaction design. *Journal on Multimodal User Interfaces*, *15*(2), 87–98. https://doi.org/10.1007/s12193-020-00332-0

WHO. (2001). *World health report 2001*. World Health Organisation.

Will, H. (2006). Psychoanalytische Kompetenzen. *Forum der Psychoanalyse*, *22*(2), 190–203.

Wöllmer, M., Al-Hames, M., Eyben, F., Schuller, B., & Rigoll, G. (2009). A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams. *Neurocomputing*, *73*(1–3), 366–380. https://doi.org/10.1016/j.neucom.2009.08.005

Wöllmer, M., Kaiser, M., Eyben, F., Schuller, B., & Rigoll, G. (2013). LSTM-modeling of continuous emotions in an audiovisual affect recognition framework. *Image and Vision Computing*, *31*(2), 153–163. https://doi.org/10.1016/j.imavis.2012.03.001

Xu, K., Chen, X., & Huang, L. (2022). Deep mind in social responses to technologies: A new approach to explaining the computers are social actors phenomena. *Computers in Human Behavior*, *134*(1), Article 107321. https://doi.org/10.1016/j.chb.2022.107321

Young, J. E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H. I., & Grinter, R. E. (2010). Evaluating human-robot interaction. *International Journal of Social Robotics*, *3*(1), 53–67. https://doi.org/10.1007/s12369-010-0081-8

Youssef, A. B., Chollet, M., Jones, H., Sabouret, N., Pelachaud, C., & Ochs, M. (2015). Towards a socially adaptive virtual agent. *Proceedings of the 15th ACM International Conference on Intelligent Virtual Agents*, 3–16. https://doi.org/10.1007/978-3-319-21996-7_1

Youssef, A. B., Sabouret, N., & Caillou, S. (2014). Subjective Evaluation of a BDI-based Theory of Mind model. *Workshop on Affect, Compagnon Artificiel, Interaction*, 120–125.

Zeng, Z., Tu, J., Pianfetti, B. M., & Huang, T. S. (2008). Audio–visual affective expression recognition through multistream fused hmm. *IEEE Transactions on Multimedia*, *10*(4), 570–577. https://doi.org/10.1109/TMM.2008.921737

Zhang, Z., Trinh, H., Chen, Q., & Bickmore, T. W. (2015). Adapting a geriatrics health counseling virtual agent for the chinese culture. *Proceedings of the 15th International Conference on Intelligent Virtual Agents*, 275–278. https://doi.org/10.1007/978-3-319-21996-7_28